UNIVERSITY OF WAIKATO

Hamilton
New Zealand

Towards a dynamic spatial microsimulation model for projecting Auckland's spatial distribution of ethnic groups

Working Paper in Economics 12/21

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

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

Jacques Poot
National Institute of Demographic and Economic Analysis (NIDEA)
and
School of Accounting, Finance and Economics
University of Waikato
and
Department of Spatial Economics
VU University Amsterdam
Email: jacques.poot@waikato.ac.nz

Abstract

In this paper we describe the development, calibration and validation of a dynamic spatial microsimulation model for projecting small area (area unit) ethnic populations in Auckland, New Zealand. The key elements of the microsimulation model are a module that projects spatial mobility (migration) within Auckland and between Auckland and the rest of the world, and a module that projects ethnic mobility. The model is developed and calibrated using 1996-2001 New Zealand Linked Census (i.e. longitudinal) data, and then projected forward to 2006. We then compare the results with the actual 2006 population. We find that in terms of indexes of overall residential sorting and ethnic diversity, our projected values are very close to the actual values. At a more disaggregated spatial scale, the model performs well in terms of the simulated normalised entropy measure of ethnic diversity for area units, but performs less well in terms of projecting residential sorting for each individual ethnic group.

Keywords

dynamic microsimulation model, ethnic identity, location transition, ethnic transition

JEL Codes

J11, R10, R15

Acknowledgements

The first author acknowledges the support provided by the University of Waikato in the form of a University of Waikato Doctoral Scholarship. The first author is also grateful for the support provided by the 2014–2020 Capturing the Diversity Dividend of Aotearoa New Zealand (CaDDANZ) programme, funded by Ministry of Business, Innovation and Employment grant UOWX1404.

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.

1. Introduction

The residential choice preferences of individuals constitute a 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. 2021a), educational qualification (Farley 1977; Denton and Massey 1988; Domina 2006), occupational status (Duncan and Duncan, 1955; Simkus 1978), and income (Fischer 2003).

Many studies have linked ethnic diversity with residential sorting. Schelling (1971) noted that all factors leading to residential sorting are interrelated. An individual looking for a place to live 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 ease of communication and trust. This results in people clustering together with others of the same ethnicity. Residential sorting may also occur in terms of other characteristics such as education. Individuals with similar level of education have similar types of jobs, resulting in similar incomes. Individuals with similar incomes may cluster together due to having similar housing affordability, and house prices and rents are spatially clustered too – as explained by the bid-rent model (Alonso, 1964; McCann 2013). Mondal et al. (2021a) captured these mutually reinforcing aspects of sorting by income, education and occupation in Auckland, New Zealand, by calculating a combined index of economic residential sorting. They found that ethnic residential sorting is much more prevalent in Auckland than economic residential 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 in time or focusses on 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, infrastructural and community facilities, as well as services such as education and healthcare (Mondal et al. 2021a; Mondal et al. 2021b). It is equally important to have knowledge about future residential sorting patterns, in order to enhance the efficiency and efficacy of 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 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, remains relatively limited (Bruch and Mare 2006). Broader understanding of changing residential sorting patterns requires an examination at different spatial levels, as different geographic scales portray different dimensions of residential sorting (Reardon et al. 2009). Urban households are likely to take spatial features apparent at different spatial scales into account when deciding on their residential location, especially 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 projecting ethnic populations more difficult. However, projecting future ethnic diversity or residential sorting at a neighbourhood scale also suffers from 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 microsimulation model (MSM) of the population of the Auckland region that captures ethnic diversity at a fine (i.e. small area) spatial scale and with the maximum feasible disaggregation of ethnic groups. The model is constructed from actual linked (i.e. longitudinal) census data from 1996-2001, then tested with census data from 2006. Consequently, the projection model can be described as a dynamic spatial MSM.

While two censuses have been held since then (in 2013 and 2018), the data from these censuses were not suitable for the development of the MSM. The 2011 census was delayed until 2013 due to a large earthquake in Christchurch. The resulting inter-censal period of seven years (instead of the usual five) has implications for the comparability of inter-censal transition probabilities. Moreover, 2018 census data were not available at the time that this project commenced, and out-of-sample projection of the population with a base year of 2013 cannot be validated until 2023 census results become available.

Hence, using 1996-2001 New Zealand Linked Census data, we construct a dynamic spatial MSM to project the ethnic distribution of the population in 2006 in Auckland, New Zealand, i.e. out of sample. We incorporate changes in ethnicity-specific population along with ethnic and spatial mobility. We test our model by comparing our simulated results to the actual 2006 census data. This work represents the first attempt to develop a dynamic spatial 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. 2021a).

We develop and run our model in Stata, which is in itself a novel approach to dynamic spatial microsimulation modelling. As Stata is available inside the secured Statistics New Zealand Datalab, we could run our model without being required to use anonymized/unit record data. which prevented any bias arising due to anonymization that might occur when a synthetic base population is used. Moreover, we were also able to use the entire 1996-2001 Auckland linked-censuses 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 2 presents detailed information about different types of MSMs and how they have been used in previous research.

Sections 3 and 4 describes the data and the methods we employ respectively, Section 5 describes the results and the testing of the MSM model, and Section 6 concludes.

2. Literature Review

Microsimulation is a method that can be used to simulate and project populations and their attributes. *Micro* refers to individual units, e.g. people (Mot 1992), households (Rogers et al. 2014), or firms (Moeckel, 2009). 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 population 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 2020b), sub-national/district level (Cameron and Cochrane 2016a), or at small area level (Cameron and Cochrane 2016b). Cameron and Poot (2019) develop and discuss ethnic population projections for New Zealand regions using a 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 paper is the first model that generates ethnic projection at a disaggregated level of ethnicities and for small spatial areas (area units).

2.1 Types of MSMs

All MSMs require micro data (Wu et al. 2011), but 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 mainly been used to evaluate tax-benefit systems (Pechmen and Okner 1974) or 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 changes in marginal and participation tax rates[1] 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 the 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 separately for 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, 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 wages. 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 any 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,

[1] Participation tax rates are the difference between current household taxes and benefits and the household taxes and benefits when the individual earnings are set to zero, divided by individual earnings (Immervoll et al. 2007).

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 a transition probability matrix for the simulations (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[2] to build their model and transition probability estimates.

Continuous-time dynamic MSMs treat time as continuous and thus are able to estimate the time at which each event occurs. In these models, individuals are assigned characteristics that can change at any time, using 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

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

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), or not at all.

MSMs can also be non-spatial or spatial in nature. *Dynamic spatial 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 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 the 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. Census data do not always provide all of 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 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 and O' Donoghue 2012).

2.2 Previous MSMs Projecting Ethnic Population Change

Dynamic MSMs have previously 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 mainly related 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. They used four broad ethnic groups (Māori, Pacific, Asian and NZ European/Other), considering them as time-invariant (i.e. each individual's ethnicity is assumed to 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. 2020a, who show similar past trends for Auckland, New Zealand's largest city). Also, changing patterns in living arrangements, such as shifting away from the nuclear family, were 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 (New Zealand European, Māori, Pacific Peoples, and Asian). He took into account the link between micro level (individual preferences) and macro-level (number of groups, group size, and proportion) elements to model and predict (until 2021) the changing ethnic residential patterns within the Greater Auckland Urban Area at both meso (territorial authorities[3]) 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). Ethnic population was projected to be consistently clustered over time 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, there would be several area units where the Pacific people were projected to be the largest group. It was also projected that in the Manukau area there would be an absolute decline in the New Zealand European population.

[3] Auckland City, Manukau, North Shore, Waitakere, and Papakura.

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 the four major ethnic groups , and 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 across the whole Auckland urban area was less 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, both globally and in New Zealand. 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. 2021b). Moreover, they focused only on four broad ethnic groups, which ignores the diversity *within* these ethnic groups (especially within the Asian and Pacific Peoples ethnic groups) (Mondal et al. 2021a). These studies also did 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; Didham 2016).

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 previous studies in New Zealand and overseas. This is necessary to capture the underlying ethnic and spatial heterogeneity in choices and characteristics (Mondal et al. 2021a;2020). Ethnic mobility is experienced by a surprisingly large proportion of people in New Zealand. 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.

3. Data

Auckland is the most ethnically diverse region in New Zealand and accounts for about one-third 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[4] (2.3 percent), and Other (1.1 percent) (Statistics New Zealand 2020a).[5] 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 comprising observations merged across a pair of censuses. Throughout this paper we use 'previous' to refer to data from the first census in each inter-censal period and 'current' for data from the following census. The link rate for individuals from the 1996 Census to 2001 Census was 69.5 percent, and for the 2001 to 2006 Census was 70.3 percent (Statistics New Zealand 2014).[6] The NZLC is the most comprehensive source of longitudinal socio-demographic information on individuals (e.g. sex, age, ethnicity, education, place of residence etc.) in New Zealand. Our analysis is based on data aggregated to the area unit level, using 2013 Auckland area unit boundaries.[7] 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.[8]

In New Zealand, ethnicity is defined as the ethnic group that people feel a sense of belonging to. It is not a measure of race, ancestry, nationality or citizenship, but a measure of cultural affiliation. Ethnicity is self-recognised, and individuals can choose up to six ethnic groups in the census.[9] Individuals are also able to choose one or more different ethnicities in each census from any they have chosen previously (Statistics New Zealand 2015).

The New Zealand Standard Classification of Ethnicity categorises ethnicity into 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 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 2021a). Thus, we use Level 2 ethnic groups in order to better capture this heterogeneity. Due to the small number of individuals

[4] Middle Eastern/Latin American/African.

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

[6] The link rate for 2013 are unavailable. A census pair '*t-5*, *t*' refers to a pair of censuses where individual records in census (*t*) are linked to those of the previous census (*t-5*). 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 (Statistics New Zealand 2014).

[7] 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.

[8] As stated in the *Disclaimer* at the start of this paper.

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

reporting as 'Not further defined' groups (among those who are European, Asian or Pacific Peoples) and 'Other' 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 group sizes are too small for some groups to develop a suitable model.

Table 1: Ethnic group classification in New Zealand

Ethnic group code (Level 1)	Ethnic Group code description	Ethnic group code (Level 2)	Ethnic Group code description (Level 2)	Ethnic group in simulation
01	European	10	European not further defined	18
		11	NZ European	1
		12	Other European	2
02	Māori	21	NZ Māori	3
03	Pacific Peoples	30	Pacific Island not further	18
		31	Samoan	4
		32	Cook Island Māori	5
		33	Tongan	6
		34	Niuean	7
		35	Tokelauan	8
		36	Fijian	9
		37	Other Pacific Island	10
04	Asian	40	Asian not further defined	18
		41	Southeast Asian	11
		42	Chinese	12
		43	Indian	13
		44	Other Asian	14
05	MELAA	51	Middle Eastern	15
		52	Latin American/Hispanic	16
		53	African	17
06	Other	61	Other ethnicity	18

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.[10] These inconsistencies particularly affect the European ethnic groups (including New Zealand European) and the Māori ethnic group. In the

[10] 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 the 1991 and 2001 Censuses. Moreover, the first two answer boxes appeared 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).

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 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.[11] 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. 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).

Our model relies on inter-censal migration flows, which generally requires that we observe the location of each individual in both censuses. That is problematic in the case of emigration (from Auckland to overseas), and mortality, as in both cases the individual is not observed in the second of each pair of linked censuses. To overcome this issue, we apportioned the number of emigrants from Auckland and number of deaths in Auckland to each area unit according to area unit population.[12] For in-migration (from overseas or elsewhere in New Zealand to Auckland) and births, we identified 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. Thus, our model accounts for both mortality and fertility although they will in practical terms be combined with out-migration and in-migration respectively.,.

The Normalised Entropy Diversity score measure[13] of residential sorting is a measure of ethnic diversity in each area unit used in the analysis (Mondal et. al. 2021a). The normalised

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

[13]The Normalised Entropy diversity measure is calculated using the formula $D_a = -\sum_{g=1}^{G} \frac{P_{ga}}{P_{..}} \ln\left(\frac{P_{ga}}{P_{..}}\right) / \ln(G)$, where P_{ga} refers to the population of group g (=1, 2,…G) in area a (= 1,2,….A), and $P_{..}$ is the total Auckland

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.

4. Methodology

In this section, we describe the construction and calibration of a dynamic spatial MSM which can be used to project the future spatial distribution of ethnic diversity in Auckland, taking both 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). As we study the geographical trends in the ethnic diversity of Auckland, our model is *spatial* in nature.

Our model is a *discrete-time* (runs in five-year time steps) *probabilistic* (uses transitional probabilities to project forward) and *dynamic* (includes time effect) spatial *MSM*. Our model is an *open 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). Additionally, the simulated population changes through spatially distributed births and deaths.

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. Although the MSM model uses data only from the linked 1996-2001 census, it simulates the whole Auckland population. It also implicitly incorporates mortality and immigration. This model can then be used to develop a *projection model* that will simulate and project predictor and predicted variables. The projection model is beyond the scope of the present paper and will be developed in follow-up work. The validation model is comprised of two modules: (1) an ethnic transition module; and (2) a locational transition module. For each of these modules, we break the population into two age groups: (1) children/adolescents (0-17 years); and (2) adults (18 years and older).

The MSM captures individual ethnic transitions as well as spatial mobility i.e. individuals making choices regarding their ethnicity and location. Figures 1 and 2 outline the theoretical framework for the ethnic transition and locational transition modules respectively.

population. To allow the calculation of D_a even in the case of there being groups who have zero members in area a at some point in time, we define $0*\ln(1/0)= \lim_{q\to 0}[\,q(\ln(1/q)] = 0$.

In practice, the ethnic transition module runs first in each time step, followed by the locational transition module.

Figures

Figure 1: Theoretical framework – Ethnic Transition

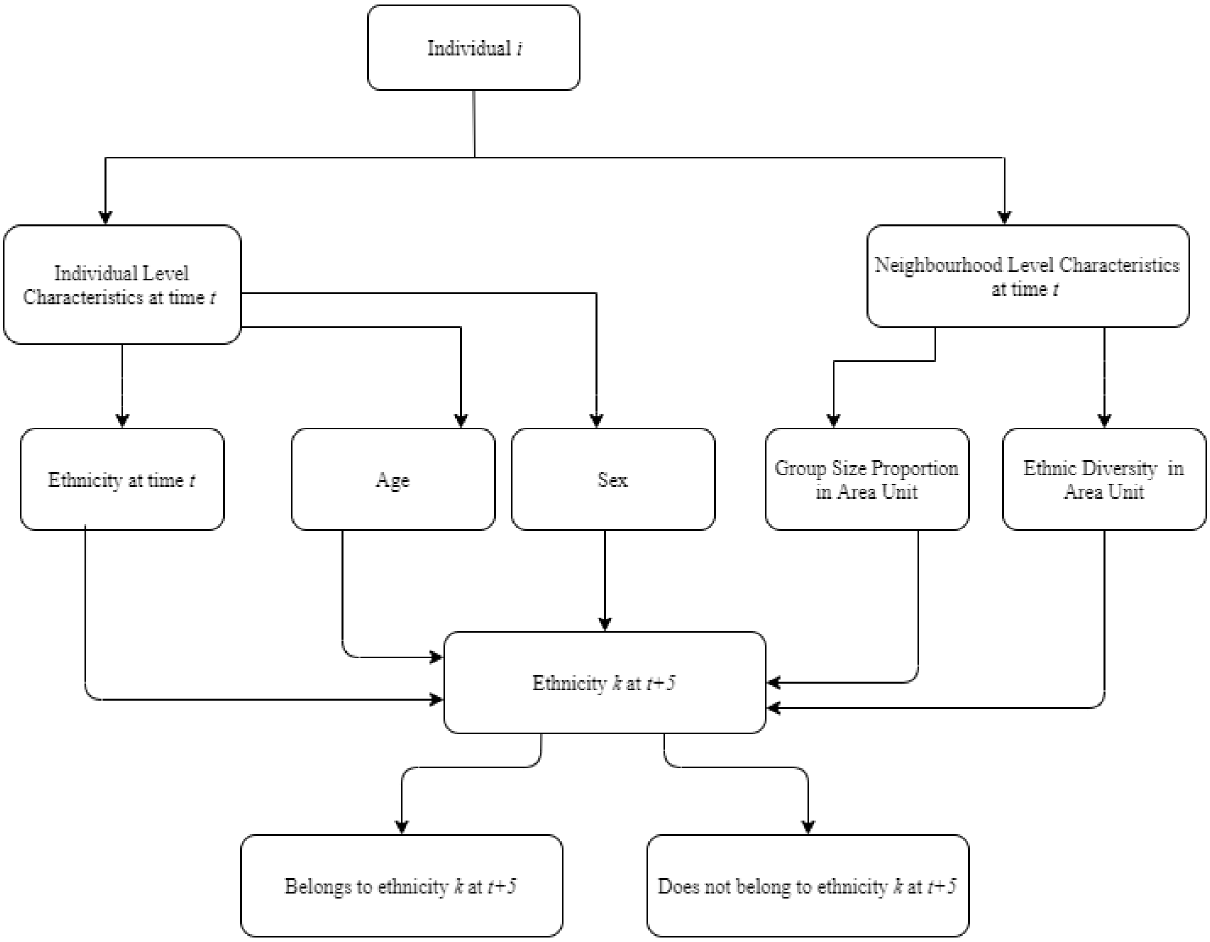

Figure 2: Theoretical framework – Location Transition

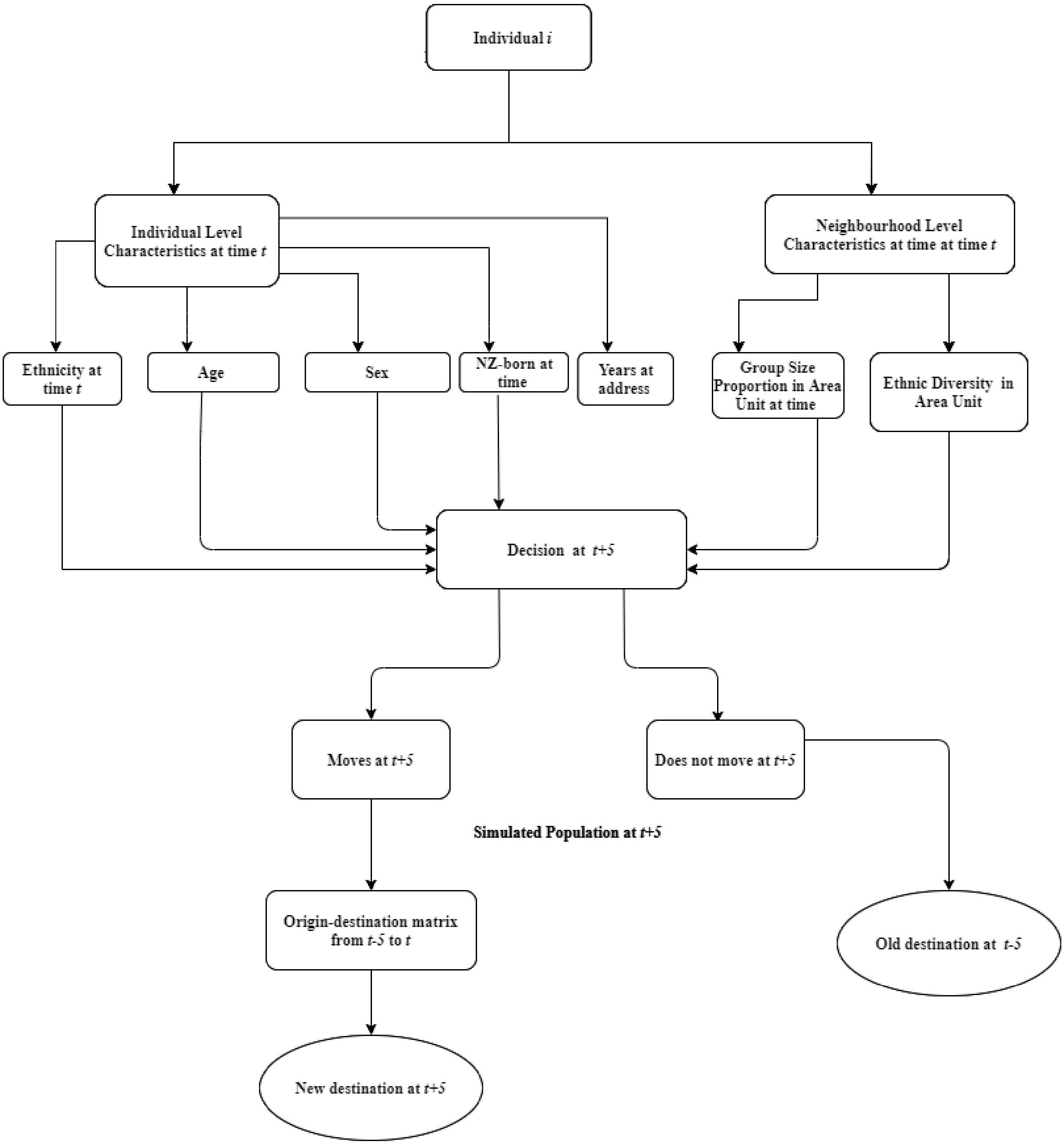

Table 2 summarises the variables used in the analysis. The ethnic transition module runs a separate logistic regression equation for each ethnicity. We take the individual's ethnic response, which is binary (1 = belongs to ethnic group *i*, 0 = otherwise), in the current census as the dependent variable. This variable represents whether or not the individual identifies with that group, 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 (Mondal et al. 2021b). Our approach 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 choice in regards to each ethnicity is given equal importance. From the logistic regression equations, we obtain the predicted probabilities of an individual belonging to ethnic group *i* in the current census.[14] 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.[15]

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 row vector of binary variables, with one variable for each of the eighteen ethnic groups (belongs to ethnic group i=1, does not belong to ethnic group i=0).

The individual-level independent variables included in our analysis for the ethnic transition module are the individual's ethnicity or ethnicities 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.

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 origin-destination matrix calculated with the intra-urban relocation data from the actual 1996-2001 linked census. Specifically, we first use logistic regression equations (one for adults and one for children) 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 to destination area units based on a column-standardised origin-destination matrix (with a zero diagonal). A different origin-destination matrix is used for each ethnic group. For individuals with multiple ethnicities, one of their ethnicities is chosen at random, and the corresponding origin-destination matrix is used. The destination for each migrant is determined again using a uniformly distributed random variable, with the origin-destination matrix used as a lookup table

[14] This is the probability that under the model the individual has ethnicity *i* in 2001.

[15] When predicted probability of the individual having ethnicity *i* is greater than the assigned random variable, the individual is assigned ethnicity *i* in 2006.

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.

As the decision to move is affected 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 locational transitional equations along with all other variables included in the ethnic transition equations.

Table 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 unit, Ethnic group size proportions in area unit.
Location Transition	Moved [16] (1=moved, 0=otherwise)	Individual	Ethnicity, Age, Sex, NZ-born, years at address
		Neighbourhood	Ethnic diversity in area unit, Ethnic group size proportions in area unit.

Note: These logit models are estimated separately for the population aged 0-17 and the population aged 18 and over.

[16] 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.

4.1 Simulation Evaluation

As mentioned earlier, in our validation model we use predictor variables from 1996 and predicted variables from 2001 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 between 2001 and 2006 in our simulated data to those in the actual 2001-2006 linked census data. Second, we compare measures of residential sorting based on the simulated data for 2006 with those based on actual 2006 census data. In our comparisons, we use different forecast error measures to estimate forecast error and bias in the model.

Though our model is parameterised on the linked 1996-2001 census, it simulates the whole Auckland population in 2006, by using out-of-sample prediction method. Since, the linked and non-linked population might differ in some ways, we employ the process of calibration to control for any 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, following the influential contribution by Theil and Finezza (1971). Entropy measures are conceptually and mathematically attractive and are the least biased by group size (Reardon and Firebaugh 2002; Modal et al. 2021b). The formulas of the measures of residential sorting and diversity used in our analysis are detailed in Table 3. In order to observe the extent to which one ethnic group is over- or under-represented in an area unit, we calculate the diversity (entropy) index (E_a) of the population in area unit *a* in terms of the given ethnic group classifications. Following Nijkamp and Poot (2015), we normalise the entropy diversity index to an evenness index I_a that varies between zero and one. 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 also use the Entropy Index of spatial sorting of group g (EIS_g), which measures 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 (see Table 3). This index varies between zero (when the group is distributed proportionally to the total population in all area units) and 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 (for which the formula is not included in Table 3), but is easier to interpret. Finally, we also calculate the normalised diversity (entropy) index I* of the whole Auckland population in terms of the given ethnic group classifications.

Table 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).

Projection error measures

Following Cameron and Cochrane (2017) and Wilson (2015), we estimate multiple measures of projection 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's Percentage Error at time *t* based on data at time *t*-5 ($PE_{t-5,t}$) is given as:

$$PE_{t-5,t} = \frac{M_t - A_t}{A_t} x100\%$$

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 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* is defined as:

$$WMAPE_{t-5,t} = \sum\nolimits_{g} \left(\left| PE^{g}_{t-5,t} \right| \frac{P_{gt}}{P_t} \right)$$

where g is the number of groups, P_{gt} is the population size of each group and P_t is size of the total Auckland population in year *t*.

The population projection error distribution is likely to be right-skewed due to 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 ($\mathrm{MedALPE}_t$), neither of which are not affected by extreme outliers. MedAPE_t is the middle of the set of ranked absolute PE_t values. MedAPE_t is a measure of precision of a 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.

4.2 Calibration Process

After performing the initial stages of model coding and running, we calibrated the model so that the simulated results based on 1996-2001 data would be as close as possible to the actual 2006 data. We assume that if the proportion of people changing their location, the proportion of people in each ethnic group, and the proportion of each ethnic group changing their location in the simulated data is close to the actual data, then the model should be able to replicate 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 more 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 added this calibration constant from the previously generated uniformly distributed random variable, thereby ensuring that the model would decrease the number of 'movers'. The model then uses this calibrated 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 many members of an ethnic group, we added a calibration constant onto the uniformly distributed random variable. We subtracted the calibration constants from the random variable if the model simulated too few members 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.[17] 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. Results

Our ultimate aim is to use our dynamic spatial MSM model to build a *projection model* that will project the population forward with minimum error. To obtain the predicted probabilities for both ethnic transition and location transition, we run logistic regression equations with clustered standard errors.[18] We find that with increase in age, having been born in New Zealand and longer duration of stay in the area unit, residential mobility decreases both for adults and children. Results show that an individual's previous census ethnicity is statistically significantly positive related to each ethnic identity. Individuals are more likely to choose the same ethnicity as they had affiliated themselves with in the previous census.

To this end, we validated the ability of the current model to replicate known 2006 census outcomes. Table 4 shows that 21 percent of people, who were in Auckland in 2001, changed their ethnicity (or at least one, in case the individual identified with multiple ethnicities) in the actual 2006 Census, whereas for the simulated 2006 Census from our model, this proportion is 22 percent. The percentage of people reporting moving from one area unit in 2001 to a different area unit in 2006 in the actual 2006 Census (40 percent) and the simulated 2006 data (42 percent) are very similar. The difference in the percentage of people moving out of Auckland between the actual (9 percent) and the simulated (6 percent) data is three percentage points.

[17] Our binary dependent variable 'moved' is dependent on every ethnicity the individual declares in the previous census.

[18] Appendix tables A1, A2, A3 and A4 show the logistic regression results.

Table 4: Comparison between simulated data and the actual 2006 Census data

Variable	Model	Actual	Difference (Model- Actual)
Ethnic change	22%	21%	1%
Location change	42%	40%	2%
Movement out of Auckland	6%	9%	-3%

Note: The table shows the difference in percentages of people, in the simulated 2006 Census data and the actual 2006 Census data.

Table 5 shows that in terms of overall ethnic residential sorting in Auckland, our simulated value for the Theil's multi-group spatial sorting index (H^*) is very close to the actual value, the difference being just -0.008. Table 5 also shows that the simulated ethnic diversity in Auckland (I^*) very closely matches the actual ethnic diversity observed in Auckland in 2006.

Table 5: Actual and simulated spatial sorting in Auckland, 2006

Measures of Residential Sorting	Model	Actual	Difference (Model-Actual)
Theil's multi-group index (H^*)	0.084	0.093	-0.008
Evenness Index (I^*)	0.654	0.656	-0.002

Note: The table shows the difference in the calculated sorting indexes based on the simulated 2006 Census data and the actual 2006 Census data.

Table 6 summarises the three forecast error measures (WMAPE, MedAPE and MedALPE) for both the Entropy Index of Segregation measure for ethnic groups and the Normalised Entropy Diversity measure for area units. The WMAPE estimates are larger than the MedAPE for 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 (-28.5 percent) value reflects that there is downward bias in the simulated values of the Entropy Index of Segregation measure (Table 6, column (A)), potentially resulting from the partial observability of all characteristics that might affect ethnic mobility. The inconsistencies in the ethnic categorisations in the 1996-2001 Census data mentioned in Section 3, which were used to parameterise the initial model, contribute to the model performance. This is demonstrated by the fact that although the simulated and the actual overall ethnic residential sorting in Auckland, are very similar (Table 5), the model does not perform well when we simulate the ethnic residential sorting for individual ethnic groups. Table 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 4.07 percent and 1.68 percent respectively.

Table 6: Model Performance

Error Measure	***EIS*** **(A)**	***I*** **(B)**
WMAPE (%)	19.34	4.07
MedAPE (%)	28.53	3.54
MedALPE (%)	-28.53	1.68

Note: *EIS* refers to Entropy Index of Segregation for ethnic group. *I* refers to Normalised Entropy diversity (area unit).

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 full census data for 2006 (linked and non-linked population). The simulated results were then compared to the actual 2006 Census data.

We have demonstrated that census data can be used to inform, calibrate and validate our model, which is generally 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. Projection errors vary with population size of a region (Smith and Shahidullah 1995; Tayman et al. 1998). Smith and Shahidullah (1995) worked on projections of total population for all census tracts in the three counties in Florida (Dade, Duval and Pinellas) and found that error measure values decline with increase in population size. Their reported MAPE[19] values ranged from 17.3 per cent to 27.6 per cent. Taymen et al. (1998), in their work on census tracts projections in San Diego County reported that in the census tracts with population size between 1,000 and 1,500 the MAPE values were as high as 56.5 per cent and 46.2 per cent respectively. Keeping in mind that the area unit population composition in

[19] Mean absolute percentage error

our work is around 1,500 the results show that our model projects the ethnic future spatial distribution in Auckland with a reasonable level of error.

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

This model is not without limitations. First, with a given set of predictor variables, logistic regression equations are used to predict the probability of a certain level of event occurring. Hence, only data from people who have been linked in the 1996-2001 NZLC could be used in estimating the logistic regression equation. However, the base population for the model is the whole Auckland population in the 2001 Census, whether linked in the 1996-2001 NZLC data or not. Thus, to the extent that unlinked and linked people differ in ways that are correlated with the transitions we estimate, that will generate bias in the results. We do not know the direction of this bias. 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 'ONFD'. 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, but at a cost of deviating the model further from the underlying real-world data from the full census. Hence we preferred to retain these ethnic groups at this stage of model development. A future extension to this work could be to separate these ethnic groups or merge them into other Level 2 groups within the same Level 1 broad ethnic group, and observe the effect on the model results. These model extensions would become easier if the 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. In that case, interregional migration will have to be modelled instead of intra-urban residential mobility.

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 (which can also proxy for income). 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.[20] Thus, we have not included education as a predictor variable 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. another important determinant of adult mobility is marital status (Speare and Goldscheider 1987; Feijten and Ham 2007). Although data on an individual's relationship status is available in the census, the data is of poor quality especially at small spatial level (Statistics New Zealand 2020c). Thus, we did not include this variable in our model.

In spite of these limitations, this paper has described the development of a modelling approach with the potential to contribute 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 and run microsimulation models.[21] Our future focus will be to use this calibrated model and the 2013-2018 NZLC data to project the future ethnic spatial distribution in Auckland forward to 2038.

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

[21] Li and O'Donoghue (2014) used Stata to present and compare six common alignment algorithms and probability transformations used in the dynamic microsimulation modelling.

References

Alonso, W. (1964). *Location and Land Use*. Harvard University Press

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

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, UK: York Publishing Services.

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

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

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, Institute for the Study of Labor, Bonn, Germany.

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

Brown, P. & A. Gray (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.

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: 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: 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*. Northmapton, UK, 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*. 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. *Soc Sci Quart*, 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). 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. *Am J Sociol*, 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

Feijten, P., & Ham, M. v. (2007). Residential mobility and migration of the divorced and separated. *Demographic Research, 17,* 623-636,638-639,641-642,645-653. Retrieved from http://ezproxy.waikato.ac.nz/login?url=https://www.proquest.com/scholarly-journals/residential-mobility-migration-divorced-separated/docview/200856110/se-2

Fischer, M. J. (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. 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). In *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. *National Centre for Social and Economic Modeling*, Canberra, Australia.

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). University of Hawaii Press, Honolulu.

Holm, E., Holme, K., Lindgren, U., & Makila, K. (2002). The SVERIGE spatial microsimulation model. Discussion paper 595, *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

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. *Tijdschr Econ Soc Ge*, 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, NATSEM, 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: 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: UNU-MERIT, Maastricht Economic and Social Research and Training Centre on Innovation and Technology*.

Li, Jinjing & ODonoghue, Cathal. *(2014).* Evaluating Binary Alignment Methods in Microsimulation Models. *Journal of Artificial Societies and Social Simulation. 17*(1), pp. 1-19. doi: 10.18564/jasss.2334

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. *Soc Forces* 67, pp. 281-315.

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. (2013). *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. *8*(2), 28–60.

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. (2021a). Group-size bias in the measurement of residential sorting. In S. Suzuki & R. Patuelli, (Eds.), *A Broad View of Regional Science: Essays in Honor of Peter Nijkamp*. Singapore: Springer Nature, Chapter 7, pp. 113-136.

Mondal, M., Cameron, M.P., & Poot, J. (2020b). Cultural and economic residential sorting of Auckland's population, 1991-2013: An entropy approach. *Journal of Geographical Systems, 23*(2): 291-330.

Mot E. S. (1992). *Survey of Microsimulation Models*. Social Security Research Committee, The Hague: VUGA.

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.

O'Sullivan (2009). Changing neighbourhoods-neighbourhoods 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).

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

Reardon, S.F., Firebaugh, G. (2002) Measures of multigroup segregation. *Sociol Methodol*, 32, pp. 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., Wohland, P. , Norman, P. & Boden, P. (2012). “Ethnic Population Projections for the UK, 2001–2051.” *Journal of Population Research* ,29 (1), 45–89. doi: 10.1007/s12546-011-9076-z

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 D (eds), *The Frontiers of Applied Demography* (Volume 9, pp.383-408). Applied Demography Series. Springer. 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., & Wolfson, M., (2000). Public pensions-Canadian analysis based on the Lifepaths generational accounting framework. 6th Nordic Seminar on Microsimulation Models, Copenhagen, Denmark.

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). World Bank Publications.

Schelling, T.C. (1971) Dynamic models of segregation. *J Math Sociol*, 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.* Academic Press, San Diego, CA.

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

Smith, S., & Shahidullah, M. (1995). An Evaluation of Population Projection Errors for Census Tracts. *Journal of the American Statistical Association, 90*(429), 64-71. doi:10.2307/2291130

Smith, K.E., Favreault, M., Ratcliffe, C.E., Butrica, B.A., Toder, E.J., & Bakija, J. (2007). *Modeling Income in the Near Term 5*. http://dx.doi.org/10.2139/ssrn.2206397

Speare, A., & Goldscheider, F. K. (1987). Effects of Marital Status Change on Residential Mobility. *Journal of Marriage and Family*, *49*(2), 455–464. https://doi.org/10.2307/352314

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

Statistics Canada. (2018). Demosim. Retrieved October 17, 2019, from https://www.statcan.gc.ca/eng/microsimulation/demosim/demosim

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

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

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

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

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

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

Statistics New Zealand. (2020a). *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).* Available from http://nzdotstat.stats.govt.nz

Statistics New Zealand (2020b). National population projections: 2020(base)–2073. Available from https://www.stats.govt.nz/information-releases/national-population-projections-2020base2073

Statistics New Zealand (2020c). 2018 Census place summaries: Auckland region. Available from https://www.stats.govt.nz/tools/2018-census-place-summaries/auckland-region#relationship-status

Tayman, J., Schafer, E., & Carter, L. (1998). The role of population size in the determination and prediction of population forecast errors: An evaluation using confidence intervals for subcounty areas. *Population Research and Policy Review*, *17*(1), 1–20. https://doi.org/10.1023/A:1005766424443

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*. North-Holland Pub. Co, Amsterdam.

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, pp. 187-194.

Uyeki, E.S. (1964) Residential distribution and stratification, 1950-1960. *Am J Sociol*,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. Paper presented at the 11th European Colloquium on Quantitative and Theoretical Geography in Durham, England, on September 3-7, 1999. 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, NIDI, The Netherlands.

Willekens, F. (2016). Migration Flows: Measurement, Analysis and Modeling. In M.J. White (Ed.), *International Handbook of Migration and Population Distribution* (pp. 225-24). *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,* 2012, pp. 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, pp. 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), Special issue, S148. doi.org/10.1016/S0140-6736(13)61402-6.pdf

White, M.J. (1986). Segregation and diversity measures in population distribution. *Popul Index*, 52, pp.198-221.

Wu, B. M., Birkin, M. H., & Rees, P. H. (2011). A Dynamic MSM With Agent Elements for Spatial Demographic Forecasting. *Social Science Computer Review*, *29*(1), 145–160. https://doi.org/10.1177/0894439310370113

Zaidi, A., & Rake, K. (2001). Dynamic microsimulation models: a review and some lessons for SAGE. The London School of Economics, Simulating Social Policy in an Ageing Society (SAGE), Discussion Paper. 2.

Zucchelli, E., Jones, A.M., & Rice, N. (2010). The evaluation of health policies through microsimulation methods. (HEDG working paper number 10/03), Department of Economics, University of York.

Appendix

Table A1: Clustered Logistic Regression of Ethnic Identity

Variables	(1) NZ European	(2) Other European	(3) NZ Māori	(4) Samoan	(5) Cook Island Māori	(6) Tongan	(7) Niuean	(8) Tokelauan	(9) Fijian	(10) Other PI	(11) SE Asian	(12) Chinese	(13) Indian	(14) Other Asian	(15) Middle Eastern	(16) Latin American	(17) African	(18) ONFD
Adults																		
NZ European	3.051***	-0.590***	-1.378***	-1.216***	-1.596***	-1.267***	-1.756***	-1.990***	-0.682***	-1.093***	-1.578***	-1.959***	-1.848***	-0.445	-0.483**	-0.923**	-0.612*	-0.314***
	(0.042)	(0.023)	(0.031)	(0.084)	(0.126)	(0.150)	(0.127)	(0.389)	(0.162)	(0.242)	(0.189)	(0.112)	(0.140)	(0.366)	(0.188)	(0.402)	(0.316)	(0.057)
Other European	1.137***	2.758***	1.684***	-1.145***	-1.151***	-1.308***	-1.149***	-1.270**	-0.386***	-0.441**	-2.393***	-2.214***	-2.285***	-1.150***	-1.130***	0.0251	0.225	1.447***
	(0.049)	(0.045)	(0.037)	(0.056)	(0.080)	(0.104)	(0.141)	(0.522)	(0.127)	(0.201)	(0.186)	(0.095)	(0.165)	(0.280)	(0.161)	(0.272)	(0.222)	(0.073)
NZ Māori	-2.463***	0.629***	5.491***	-0.564***	-0.594***	-0.418***	-0.901***	-1.178**	0.334*	-0.197	-0.750***	-0.384***	-0.889***	0.375	-0.0392	0.305	-0.405	0.595***
	(0.049)	(0.046)	(0.050)	(0.101)	(0.178)	(0.151)	(0.142)	(0.469)	(0.189)	(0.320)	(0.273)	(0.098)	(0.148)	(0.322)	(0.279)	(0.567)	(0.504)	(0.115)
Samoan	-2.420***	-0.360***	-1.493***	7.605***	-1.282***	-1.490***	-1.123***	0.231	-0.449**	0.289	-4.051***	-1.691***	-2.482***	-1.283**	-2.128***	-1.477*	-1.794**	-0.904***
	(0.049)	(0.078)	(0.078)	(0.060)	(0.159)	(0.177)	(0.193)	(0.326)	(0.201)	(0.316)	(0.503)	(0.140)	(0.257)	(0.588)	(0.493)	(0.778)	(0.768)	(0.167)
Cook Island Māori	-2.328***	-0.578***	-1.633***	-0.912***	7.705***	-1.129***	-1.250***	-0.964*	-0.530	1.402***	-3.397***	-1.881***	-2.357***	-0.247	-1.994**	-1.485	-0.663	-0.732***
	(0.067)	(0.084)	(0.106)	(0.174)	(0.075)	(0.207)	(0.218)	(0.525)	(0.457)	(0.347)	(0.607)	(0.268)	(0.339)	(0.455)	(0.793)	(1.112)	(0.805)	(0.204)
Tongan	-2.057***	-0.596***	-1.119***	-1.099***	-1.377***	7.669***	-1.762***	-1.227***	-0.367	0.113	-3.804***	-2.010***	-2.249***	-0.813	-0.970*	-2.063*	-0.244	-1.030***
	(0.076)	(0.102)	(0.104)	(0.157)	(0.245)	(0.086)	(0.203)	(0.429)	(0.256)	(0.432)	(0.600)	(0.301)	(0.318)	(0.587)	(0.562)	(1.224)	(0.510)	(0.251)
Niuean	-2.356***	-0.478***	-1.072***	-1.230***	-0.940***	-1.403***	8.293***	-0.683	-0.615*	-0.352	-2.913***	-1.326***	-2.249***	0.135	-1.196	-0.382	-0.399	-1.485***
	(0.085)	(0.122)	(0.095)	(0.234)	(0.256)	(0.208)	(0.099)	(0.681)	(0.332)	(0.614)	(0.536)	(0.271)	(0.542)	(0.557)	(0.770)	(0.591)	(0.716)	(0.388)
Tokelauan	-2.729***	-0.390	-1.207**	-1.064***	-1.731	-0.830	-0.144	9.050***	-2.062*	0.238		-2.032**		0.461				
	(0.383)	(0.417)	(0.533)	(0.408)	(1.698)	(0.827)	(0.775)	(0.261)	(1.128)	(0.909)		(0.860)		(1.649)				
Fijian	-0.793***	-0.228*	-0.965***	-0.627*	-0.115	-0.353	-0.0513	1.693***	7.523***	1.933***	-3.170***	-0.696**	0.743***	0.394			0.311	1.993***
	(0.100)	(0.130)	(0.182)	(0.332)	(0.331)	(0.322)	(0.405)	(0.647)	(0.150)	(0.380)	(1.041)	(0.341)	(0.253)	(0.693)			(1.034)	(0.143)
Other PI	-0.904***	0.863***	-1.161***	-0.532	1.187***	-0.920	0.218	-0.267	3.052***	7.094***	-0.648	-0.282	-1.172	2.053***			0.854	1.140***
	(0.173)	(0.163)	(0.341)	(0.428)	(0.449)	(0.639)	(0.360)	(0.876)	(0.392)	(0.187)	(0.684)	(0.548)	(0.772)	(0.596)			(1.496)	(0.307)
SE Asian	-2.118***	-0.832***	-0.550**	-2.456***	-2.766***	-3.601***			-0.907	-0.159	5.708***	-0.183	-1.745***	0.210	-1.295**		0.236	1.205***
	(0.099)	(0.152)	(0.216)	(0.417)	(0.987)	(1.016)			(0.577)	(0.625)	(0.180)	(0.302)	(0.314)	(0.532)	(0.575)		(0.720)	(0.137)
Chinese	-2.756***	-1.014***	-0.649***	-0.784***	-1.436***	-2.178***	-1.965***	-1.696**	-0.489**	-0.845**	-1.226***	6.807***	-2.620***	-1.191*	-2.705***		-1.336**	-0.270*
	(0.106)	(0.099)	(0.080)	(0.085)	(0.191)	(0.299)	(0.314)	(0.847)	(0.248)	(0.377)	(0.171)	(0.088)	(0.255)	(0.638)	(0.764)		(0.625)	(0.153)
Indian	-3.029***	-1.064***	-1.131***	-2.172***	-2.041***	-3.175***	-3.374***	-1.349*	-1.559***	-0.0875	-2.518***	-2.396***	7.760***	-0.660	-1.024**		0.259	-0.620***
	(0.095)	(0.130)	(0.117)	(0.236)	(0.279)	(0.429)	(0.502)	(0.760)	(0.337)	(0.439)	(0.320)	(0.276)	(0.095)	(0.717)	(0.443)		(0.506)	(0.178)
Other Asian	-2.377***	-0.614***	-0.742	-1.769***	-1.675***	-1.209***	-1.993***	-0.109	-0.561		-2.513***	-2.125***	-0.274	9.325***				0.115
	(0.146)	(0.217)	(0.525)	(0.479)	(0.461)	(0.440)	(0.517)	(0.629)	(0.632)		(0.655)	(0.487)	(0.302)	(0.239)				(0.310)
Middle Eastern	-1.096***	-0.0886	-1.523***	-1.215**	-0.851		-1.184**	1.345	0.404	1.761**	-2.202**	-3.011***	-1.474**	0.436	7.402***	1.069	1.182	1.120***
	(0.150)	(0.170)	(0.337)	(0.599)	(0.627)		(0.529)	(1.189)	(1.017)	(0.798)	(0.917)	(0.767)	(0.737)	(0.853)	(0.162)	(1.005)	(1.076)	(0.220)
Latin American	-1.332***	1.286***	-0.082	-3.162***		-3.882***	-0.353		-0.599*		-1.409**	-0.841	-0.652			8.272***		1.332***
	(0.201)	(0.245)	(0.555)	(0.541)		(0.864)	(0.510)		(0.337)		(0.650)	(0.740)	(0.673)			(0.305)		(0.312)
African	-0.317*	-0.276	-0.741***	-0.615	-0.155	-1.833**	-0.997*		0.703			-1.750*	0.0380			0.0239	7.596***	1.998***
	(0.168)	(0.239)	(0.253)	(0.646)	(0.817)	(0.803)	(0.606)		(1.034)			(0.930)	(0.896)			(0.634)	(0.221)	(0.223)
ONFD	-0.389***	0.603***	-0.490*	0.114	-2.203*	-1.122	-1.353**		0.254	0.957	2.533***	-0.163	0.805*	1.611**	1.271**	2.294***	0.482	3.413***
	(0.128)	(0.149)	(0.295)	(0.359)	(1.133)	(0.774)	(0.688)		(0.379)	(0.725)	(0.282)	(0.478)	(0.435)	(0.630)	(0.603)	(0.677)	(0.707)	(0.132)
Children																		
NZ European	3.682***	-0.033	-0.650***	-1.008***	-0.978***	-1.124***	-0.788***	-0.920*	-0.757***	-0.468*	-0.793***	-1.443***	-1.692***	-0.762**	-0.016	-0.389	0.099	0.152
	(0.060)	(0.085)	(0.071)	(0.103)	(0.127)	(0.147)	(0.180)	(0.496)	(0.262)	(0.273)	(0.200)	(0.163)	(0.233)	(0.372)	(0.468)	(0.587)	(0.700)	(0.200)
Other European	1.578***	2.789***	1.793***	-0.394***	-0.479***	-0.302	-0.430*	0.651	-0.032	0.479	-0.376	-1.224***	-1.472***	0.578	-0.816	0.064	0.736	1.502***
	(0.065)	(0.063)	(0.056)	(0.117)	(0.137)	(0.202)	(0.259)	(0.696)	(0.347)	(0.390)	(0.304)	(0.221)	(0.356)	(0.425)	(0.530)	(0.639)	(0.499)	(0.134)
NZ Māori	-0.755***	0.921***	5.040***	-0.343***	0.246**	-0.493***	0.115	0.493	0.272	1.078***	-0.917***	-0.201	-0.657***	0.990**	0.475	-0.235	-1.637*	0.389**
	(0.085)	(0.0950)	(0.070)	(0.126)	(0.119)	(0.147)	(0.179)	(0.509)	(0.249)	(0.277)	(0.315)	(0.195)	(0.252)	(0.403)	(0.662)	(0.847)	(0.947)	(0.180)
Samoan	-1.410***	0.470***	-1.054***	6.672***	-0.568***	-0.327	-0.430	1.877***	-0.471	1.007***	-1.329**	-0.876***	-1.330***	-1.364	-1.130	-1.560**	-0.804	0.100
	(0.060)	(0.099)	(0.074)	(0.087)	(0.192)	(0.208)	(0.265)	(0.501)	(0.321)	(0.375)	(0.550)	(0.225)	(0.371)	(0.965)	(1.071)	(0.721)	(0.588)	(0.295)
Cook Island Māori	-1.155***	0.693***	-1.073***	-0.377**	6.704***	-0.532**	-0.058	-0.104	-0.136	1.712***	-1.054	-1.057***	-1.634***	-0.270	0.591	0.483	0.796	0.843***
	(0.074)	(0.151)	(0.154)	(0.173)	(0.116)	(0.208)	(0.225)	(0.721)	(0.330)	(0.338)	(0.662)	(0.391)	(0.511)	(0.772)	(0.714)	(1.033)	(0.907)	(0.250)
Tongan	-1.428***	0.477***	-1.136***	-0.789***	-0.657***	6.742***	-0.457	0.796	-0.015	0.024	-1.487***	-1.099***	-2.410***	0.101	0.981			0.187
	(0.112)	(0.181)	(0.124)	(0.204)	(0.247)	(0.126)	(0.419)	(0.633)	(0.524)	(0.544)	(0.555)	(0.397)	(0.620)	(0.711)	(0.738)			(0.415)
Niuean	-0.929***	0.479**	-0.430***	-0.484*	-0.211	-0.031	7.456***	0.062	0.689*	-0.152	-1.936**	-0.221	-2.176***		0.201	0.684		0.536
	(0.118)	(0.188)	(0.100)	(0.256)	(0.236)	(0.312)	(0.139)	(0.922)	(0.418)	(0.552)	(0.752)	(0.385)	(0.613)		(1.127)	(0.632)		(0.397)
Tokelauan	-0.860*	0.564	-1.144***	-0.398		-1.937	0.798	8.803***		2.761***		0.641						

	(0.499)	(0.771)	(0.357)	(0.968)		(1.199)	(0.744)	(0.634)		(0.679)		(0.819)						
Fijian	-0.595***	0.866***	-0.557***	-0.550	-0.115	0.431	0.231		7.149***	2.054***		0.470	0.972**	1.832*			0.307	2.130***
	(0.200)	(0.264)	(0.199)	(0.456)	(0.439)	(0.660)	(0.625)		(0.223)	(0.701)		(0.778)	(0.478)	(1.084)			(0.902)	(0.365)
Other PI	-0.105	1.128***	-1.166***	-0.896	-0.020	-0.606	-0.419	-0.538	2.604**	6.390***		0.978	-0.536	1.779**				1.939***
	(0.327)	(0.391)	(0.421)	(1.061)	(0.503)	(1.513)	(0.322)	(1.185)	(1.016)	(0.363)		(1.330)	(0.602)	(0.889)				(0.551)
SE Asian	-1.509***	0.035	-1.359***	-3.143***	-1.252*	-0.837*	-1.111				6.623***	0.746	-0.937**	0.240	1.054			2.863***
	(0.145)	(0.256)	(0.294)	(0.934)	(0.642)	(0.494)	(0.800)				(0.194)	(0.468)	(0.429)	(1.057)	(0.859)			(0.232)
Chinese	-1.741***	0.350**	-0.801***	-0.780***	-0.811***	-1.740***	-0.536	-0.139	-0.618	-0.0403	0.870*	6.399***	-1.830***	-0.385	1.105	-0.0796		1.445***
	(0.118)	(0.165)	(0.124)	(0.172)	(0.252)	(0.433)	(0.372)	(0.856)	(0.448)	(0.541)	(0.446)	(0.129)	(0.619)	(1.242)	(1.000)	(0.915)		(0.243)
Indian	-2.091***	-0.099	-1.288***	-1.220***	-1.981***	-1.762***	-0.631*	0.443	-0.371	-0.410	-0.181	-1.679***	7.827***	0.335	-0.137	0.641	1.196	1.034***
	(0.122)	(0.239)	(0.139)	(0.337)	(0.410)	(0.462)	(0.358)	(1.180)	(0.651)	(0.810)	(0.491)	(0.604)	(0.170)	(0.843)	(1.078)	(1.343)	(1.031)	(0.380)
Other Asian	-1.563***	-0.067	-0.793**	-1.658	-0.841***	0.432	0.386					-1.765***	-2.185	8.719***				1.274***
	(0.266)	(0.576)	(0.361)	(1.627)	(0.326)	(0.661)	(1.081)					(0.457)	(1.959)	(0.373)				(0.494)
Middle Eastern	-0.310	-0.112	-2.215***	0.994	-0.023	-0.023					-1.023	1.006			8.182***		2.912**	
	(0.272)	(0.571)	(0.617)	(1.011)	(1.041)	(1.016)					(0.915)	(1.101)			(0.580)		(1.234)	
Latin American	-0.147	2.126***	-2.323**	-2.250**	0.619		-0.298					-0.118	0.841			8.550***		2.093**
	(0.418)	(0.592)	(1.093)	(1.050)	(1.023)		(0.591)					(0.668)	(1.091)			(0.543)		(1.051)
African	-0.734	1.071**	-1.097*	-0.667	1.057				1.300***								8.460***	1.478*
	(0.554)	(0.537)	(0.576)	(0.485)	(0.927)				(0.484)								(0.594)	(0.841)
ONFD	0.290	0.586	-0.369	1.081**	-1.591		1.114		1.011***	1.098	3.092***	0.502	-0.802	0.395	1.089		1.094	4.474***
	(0.266)	(0.410)	(0.458)	(0.443)	(1.001)		(0.975)		(0.357)	(0.751)	(0.469)	(0.948)	(2.145)	(0.737)	(2.231)		(0.690)	(0.284)

Notes:

The table reports logistic regression coefficients.

We have combined the ‘Other’ and ‘not further defined’ (among those who are European, Asian or Pacific Islanders) ethnic groups into one group ‘ONFD’. Thus, our analysis includes eighteen Level 2 ethnic groups instead of twenty-one (see Table 1).

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

Clustered Standard errors in parenthesis

Tables A1, A2 and A3 are reporting results from the same regression. We have broken down the results into different tables according to blocks of explanatory variables for easy readability.

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

The number of observations and goodness of fit measures are given at the bottom of Table A3.

Table A2: Clustered Logistic Regression of Ethnic Transition-Effect of Individual-Level Variable

Variables	(1) NZ European	(2) Other European	(3) NZ Māori	(4) Samoan	(5) Cook Island Māori	(6) Tongan	(7) Niuean	(8) Tokelauan	(9) Fijian	(10) Other PI	(11) SE Asian	(12) Chinese	(13) Indian	(14) Other Asian	(15) Middle Eastern	(16) Latin American	(17) African	(18) ONFD
Adults																		
Sex	-0.027**	-0.039***	-0.115***	-0.018	-0.148**	0.036	0.031	0.082	-0.057	-0.390***	-0.211**	-0.033	0.151**	0.136	0.014	-0.171	0.121	-0.157***
	(0.014)	(0.014)	(0.021)	(0.050)	(0.074)	(0.071)	(0.082)	(0.219)	(0.081)	(0.105)	(0.083)	(0.049)	(0.066)	(0.158)	(0.099)	(0.205)	(0.156)	(0.033)
Age	0.013***	-0.012***	-0.023***	-0.036***	-0.035***	-0.039***	-0.029***	-0.039***	-0.024***	-0.027***	-0.035***	-0.025***	-0.034***	-0.033***	-0.008**	-0.039***	-0.028***	-0.016***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.004)	(0.005)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)	(0.010)	(0.007)	(0.001)
NZ Born	2.288***	-1.552***	3.256***	-1.181***	-0.627***	-1.501***	-0.501***	-0.270	-1.066***	-0.517***	-2.924***	-1.005***	-1.776***	-2.141***	-1.484***	-2.438***	-1.187***	-1.968***
	(0.038)	(0.035)	(0.052)	(0.077)	(0.119)	(0.132)	(0.122)	(0.290)	(0.136)	(0.180)	(0.142)	(0.087)	(0.131)	(0.332)	(0.184)	(0.371)	(0.280)	(0.076)
Children																		
Sex	-0.016	-0.121**	-0.064**	-0.024	-0.18**	-0.022	-0.030	-0.147	0.146	-0.326	-0.039	0.118	-0.259**	-0.019	-0.560**	-0.186	-0.247	0.111
	(0.029)	(0.050)	(0.030)	(0.069)	(0.089)	(0.086)	(0.117)	(0.256)	(0.180)	(0.213)	(0.133)	(0.089)	(0.130)	(0.246)	(0.236)	(0.423)	(0.392)	(0.115)
Age	-0.007	-0.018	-0.032***	-0.021	0.012	-0.062***	-0.013	0.014	0.002	-0.089*	-0.074**	-0.005	-0.028	-0.078	0.028	0.104	-0.043	-0.070***
	(0.007)	(0.0113)	(0.008)	(0.014)	(0.017)	(0.018)	(0.026)	(0.061)	(0.040)	(0.046)	(0.030)	(0.019)	(0.029)	(0.056)	(0.058)	(0.092)	(0.083)	(0.024)
NZ Born	1.380***	-1.078***	1.043***	-0.060	0.071	-0.282*	0.093	-0.115	-0.149	-1.058***	-0.984***	-0.527***	-0.599**	-2.001***	-0.926***	-1.089*	-0.308	-0.907***
	(0.068)	(0.072)	(0.085)	(0.136)	(0.175)	(0.154)	(0.232)	(0.664)	(0.273)	(0.285)	(0.237)	(0.150)	(0.237)	(0.372)	(0.307)	(0.569)	(0.671)	(0.144)

Notes:

The table reports logistic regression coefficients.

We have combined the 'Other' and 'not further defined' (among those who are European, Asian or Pacific Islanders) ethnic groups into one group 'ONFD'. Thus, our analysis includes eighteen Level 2 ethnic groups instead of twenty-one (see Table 1).

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

Clustered Standard errors in parenthesis

Tables A1, A2 and A3 are reporting results from the same regression. We have broken down the results into different tables according to blocks of explanatory variables for easy readability.

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

The number of observations and goodness of fit measures are given at the bottom of Table A3.

Table A3: Clustered Logistic Regression of Ethnic Transition-Effect of Neighbourhood-Level Characteristics

Variables	(1) NZ European	(2) Other European	(3) NZ Māori	(4) Samoan	(5) Cook Island Māori	(6) Tongan	(7) Niuean	(8) Tokelauan	(9) Fijian	(10) Other PI	(11) SE Asian	(12) Chinese	(13) Indian	(14) Other Asian	(15) Middle Eastern	(16) Latin American	(17) African	(18) ONFD
Adults																		
Entropy	-0.486	0.974***	0.595*	3.348***	2.006	1.720	0.493	4.857	-0.443	-2.507**	1.158	1.319*	1.439	4.316*	4.983***	-0.512	-1.758	-0.033
	(0.319)	(0.293)	(0.362)	(1.015)	(1.541)	(1.107)	(1.260)	(4.075)	(0.977)	(1.263)	(1.320)	(0.748)	(1.234)	(2.255)	(1.889)	(2.421)	(3.184)	(0.703)
NZ European Gr	-0.009***	-0.030***	-0.058***	-0.042***	-0.041***	-0.039***	-0.046***	-0.092***	-0.068***	-0.057***	-0.035***	-0.028***	-0.020***	-0.061***	-0.060***	-0.043***	-0.084***	-0.041***
	(0.001)	(0.001)	(0.002)	(0.004)	(0.006)	(0.005)	(0.006)	(0.022)	(0.006)	(0.008)	(0.006)	(0.003)	(0.006)	(0.010)	(0.008)	(0.012)	(0.013)	(0.003)
Other European Gr	-0.003	-0.029***	-0.051***	-0.032***	-0.048***	-0.047***	-0.055***	-0.006	0.003	-0.009	-0.003	-0.034***	-0.068***	-0.065***	-0.076***	-0.101**	0.024	-0.022***
	(0.003)	(0.004)	(0.005)	(0.009)	(0.012)	(0.014)	(0.015)	(0.050)	(0.016)	(0.021)	(0.016)	(0.010)	(0.013)	(0.025)	(0.018)	(0.046)	(0.028)	(0.007)
NZ Māori Gr	-0.025***	-0.025***	0.007*	-0.019**	-0.017	-0.026**	-0.021	-0.091**	-0.034***	0.0002	-0.010	-0.044***	-0.021*	-0.062**	-0.113***	-0.069**	-0.031	-0.013*
	(0.003)	(0.003)	(0.004)	(0.009)	(0.014)	(0.012)	(0.013)	(0.038)	(0.011)	(0.016)	(0.013)	(0.010)	(0.011)	(0.025)	(0.019)	(0.029)	(0.033)	(0.007)
Samoan Gr	-0.013**	-0.030***	-0.045***	-0.012	-0.029*	-0.023*	-0.045***	-0.053	-0.048***	-0.045**	-0.033	-0.010	-0.018	-0.093***	-0.093*	-0.058	-0.0326	-0.029***
	(0.005)	(0.005)	(0.005)	(0.011)	(0.015)	(0.013)	(0.014)	(0.040)	(0.018)	(0.021)	(0.021)	(0.011)	(0.019)	(0.033)	(0.049)	(0.043)	(0.046)	(0.011)
Cook Island Māori Gr	-0.044***	-0.054***	-0.093***	-0.071***	-0.060**	-0.091***	-0.034	-0.090	-0.056*	-0.176***	-0.025	-0.097***	-0.081**	-0.199***	0.0189	0.0241	-0.078	-0.080***
	(0.009)	(0.011)	(0.012)	(0.016)	(0.024)	(0.025)	(0.030)	(0.070)	(0.029)	(0.047)	(0.041)	(0.023)	(0.034)	(0.072)	(0.056)	(0.058)	(0.069)	(0.020)
Tongan Gr	-0.020***	-0.047***	-0.049***	-0.076***	-0.006	0.006	-0.107***	-0.153*	-0.064**	0.038	-0.121***	-0.023	-0.040	-0.024	-0.204***	-0.224***	-0.010	-0.046**
	(0.007)	(0.009)	(0.010)	(0.021)	(0.026)	(0.032)	(0.040)	(0.085)	(0.031)	(0.047)	(0.037)	(0.025)	(0.037)	(0.054)	(0.058)	(0.087)	(0.065)	(0.019)
Niuean Gr	-0.005	0.005	0.002	-0.050	-0.043	0.019	0.092**	-0.023	-0.063	0.050	0.134**	-0.038	-0.025	0.002	0.010	0.091	-0.031	-0.019
	(0.012)	(0.013)	(0.019)	(0.032)	(0.039)	(0.034)	(0.042)	(0.132)	(0.054)	(0.069)	(0.059)	(0.040)	(0.060)	(0.105)	(0.087)	(0.140)	(0.101)	(0.030)
Tokelauan Gr	-0.017	0.033	0.026	0.380**	0.072	0.124	0.143	1.091	0.615*	0.756**	-0.352	0.129	0.073	0.207	-0.525	-0.891	0.211	0.150
	(0.064)	(0.078)	(0.094)	(0.160)	(0.198)	(0.193)	(0.293)	(0.773)	(0.318)	(0.368)	(0.394)	(0.218)	(0.344)	(0.692)	(0.544)	(0.841)	(0.616)	(0.166)
Fijian Gr	-0.022	0.140***	0.105**	0.205**	0.081	0.240**	0.184	-0.309	0.549***	0.029	0.039	0.042	0.176	0.210	0.118	1.009**	0.111	0.025
	(0.0295)	(0.0356)	(0.0466)	(0.0805)	(0.111)	(0.110)	(0.127)	(0.362)	(0.179)	(0.194)	(0.184)	(0.113)	(0.137)	(0.290)	(0.185)	(0.417)	(0.287)	(0.0805)
Other PI Gr	-0.024	0.209***	0.169**	-0.014	0.182	0.197	0.127	-0.216	-0.085	0.826***	-0.322	-0.059	0.080	0.409	0.371	0.990	-0.478	0.0127
	(0.052)	(0.065)	(0.077)	(0.152)	(0.156)	(0.208)	(0.265)	(0.645)	(0.308)	(0.320)	(0.327)	(0.203)	(0.258)	(0.440)	(0.416)	(0.653)	(0.459)	(0.122)
SE Asian Gr	0.008	-0.007	0.034	-0.091	0.023	0.066	0.001	0.113	-0.033	0.154	0.295***	-0.045	-0.067	-0.011	-0.008	0.380**	-0.281	-0.106**
	(0.017)	(0.021)	(0.031)	(0.057)	(0.079)	(0.069)	(0.086)	(0.178)	(0.110)	(0.119)	(0.082)	(0.070)	(0.072)	(0.159)	(0.127)	(0.156)	(0.186)	(0.048)
Chinese Gr	-0.018***	-0.042***	-0.076***	-0.072***	-0.057**	-0.0388	-0.085***	-0.137	-0.029	-0.083**	0.0002	0.042***	-0.036	-0.047	-0.047	0.057	-0.062	-0.026**
	(0.005)	(0.006)	(0.010)	(0.022)	(0.027)	(0.029)	(0.027)	(0.099)	(0.030)	(0.037)	(0.031)	(0.015)	(0.024)	(0.047)	(0.033)	(0.076)	(0.061)	(0.013)
Indian Gr	-0.036***	-0.025**	-0.051***	-0.021	-0.055	-0.024	0.062**	-0.020	-0.0001	-0.051	-0.022	-0.050**	0.067**	-0.179***	-0.176***	-0.123	0.058	-0.007
	(0.007)	(0.010)	(0.010)	(0.020)	(0.036)	(0.031)	(0.029)	(0.090)	(0.032)	(0.050)	(0.038)	(0.024)	(0.029)	(0.060)	(0.057)	(0.089)	(0.066)	(0.018)
Other Asian Gr	-0.045	-0.015	-0.169***	0.276***	0.160	0.214	0.298*	-0.722	-0.031	0.059	-0.341*	0.011	0.090	0.232	0.110	-0.159	0.819***	0.017
	(0.034)	(0.038)	(0.061)	(0.099)	(0.148)	(0.131)	(0.179)	(0.516)	(0.180)	(0.267)	(0.204)	(0.109)	(0.131)	(0.266)	(0.242)	(0.398)	(0.302)	(0.077)
Middle Eastern Gr	-0.155***	0.230***	0.040	0.090	0.056	0.168	0.049	0.101	-0.390	0.101	0.107	-0.116	0.142	0.527***	0.916***	0.198	0.057	0.041
	(0.048)	(0.049)	(0.066)	(0.101)	(0.102)	(0.112)	(0.233)	(0.222)	(0.265)	(0.129)	(0.178)	(0.173)	(0.184)	(0.175)	(0.153)	(0.527)	(0.359)	(0.096)
Latin American Gr	0.161*	-0.443***	-0.126	0.050	0.293	-0.130	-0.717	-0.966	-0.055	-0.623	-0.126	0.211	0.029	-0.738	0.872	1.827*	1.137	0.132
	(0.092)	(0.114)	(0.119)	(0.313)	(0.348)	(0.421)	(0.483)	(1.091)	(0.499)	(0.685)	(0.486)	(0.337)	(0.495)	(0.852)	(0.816)	(0.939)	(0.837)	(0.196)
African Gr	0.260***	-0.107	-0.026	0.002	0.544	0.441	0.558	2.148***	1.460***	-0.472	-0.245	0.005	-0.538	0.762	-0.355	-0.505	0.548	0.047
	(0.081)	(0.114)	(0.134)	(0.268)	(0.346)	(0.373)	(0.446)	(0.781)	(0.426)	(0.438)	(0.463)	(0.311)	(0.350)	(0.793)	(0.510)	(1.061)	(0.541)	(0.173)
ONFD Gr	0.063	-0.061	-0.044	0.016	-0.021	0.248	0.541*	0.897	0.169	0.391	-0.460	0.004	0.023	-0.149	-0.507	-0.737	0.699	0.212
	(0.060)	(0.074)	(0.089)	(0.179)	(0.234)	(0.229)	(0.315)	(0.597)	(0.397)	(0.386)	(0.325)	(0.191)	(0.239)	(0.650)	(0.405)	(1.070)	(0.590)	(0.130)
Observations	403,200	403,200	403,200	403,200	402,900	402,600	400,500	399,300	403,200	401,800	402,600	403,100	402,800	402,700	399,400	380,500	401,700	402,800
Pseudo R-squared	0.652	0.291	0.695	0.852	0.831	0.834	0.842	0.791	0.464	0.610	0.701	0.827	0.869	0.815	0.227	0.202	0.211	0.193
Children																		
Entropy	-2.225***	-1.102	0.899	1.949	3.464***	3.458*	1.444	-6.839***	-1.419	-1.260	1.346	0.677	7.174***	6.333	11.79**	6.622	-3.518	4.387**
	(0.711)	(0.707)	(0.555)	(1.413)	(1.283)	(1.990)	(1.954)	(1.732)	(2.569)	(2.854)	(2.501)	(1.212)	(2.354)	(4.524)	(5.423)	(5.322)	(10.54)	(1.891)
NZ European Gr	-0.005**	-0.030***	-0.047***	-0.053***	-0.0531***	-0.041***	-0.060***	-0.068***	-0.062***	-0.040**	-0.054***	-0.001	-0.054***	-0.041**	-0.072***	-0.073***	-0.100***	-0.059***
	(0.003)	(0.003)	(0.003)	(0.006)	(0.005)	(0.008)	(0.010)	(0.020)	(0.013)	(0.016)	(0.012)	(0.031)	(0.007)	(0.020)	(0.024)	(0.022)	(0.036)	(0.008)
Other European Gr	0.018***	-0.033***	-0.024***	-0.025	-0.075***	-0.078***	-0.073***	-0.010	-0.055	-0.029	-0.020	-0.0003	-0.069**	-0.179***	-0.163***	-0.151**	0.050	-0.033*
	(0.007)	(0.010)	(0.007)	(0.015)	(0.016)	(0.020)	(0.025)	(0.075)	(0.039)	(0.031)	(0.025)	(0.031)	(0.028)	(0.048)	(0.051)	(0.060)	(0.072)	(0.017)
NZ Māori Gr	-0.020***	-0.019**	0.029***	-0.005	-0.051***	-0.058***	-0.041**	0.061**	-0.021	-0.032	0.001	-0.007	-0.068***	-0.109**	-0.172***	-0.126	0.0185	-0.052***
	(0.005)	(0.009)	(0.006)	(0.012)	(0.012)	(0.018)	(0.020)	(0.030)	(0.022)	(0.032)	(0.026)	(0.024)	(0.025)	(0.044)	(0.044)	(0.079)	(0.086)	(0.019)
Samoan Gr	-0.005	-0.021	-0.029***	0.005	-0.042***	-0.033**	-0.048**	0.036	-0.054	-0.047	-0.041	-0.004	-0.031	0.011	-0.229**	0.117	0.039	-0.098***
	(0.010)	(0.015)	(0.008)	(0.011)	(0.015)	(0.016)	(0.022)	(0.052)	(0.047)	(0.055)	(0.027)	(0.032)	(0.035)	(0.071)	(0.108)	(0.088)	(0.080)	(0.031)
Cook Island Māori Gr	-0.026	-0.061**	-0.112***	-0.132***	-0.061*	-0.072*	-0.097**	-0.275**	-0.164**	-0.108	-0.144***	-0.001	-0.082	-0.211**	-0.119	-0.445**	-0.293	-0.106*
	(0.017)	(0.024)	(0.018)	(0.024)	(0.034)	(0.041)	(0.042)	(0.131)	(0.081)	(0.079)	(0.054)	(0.055)	(0.056)	(0.092)	(0.138)	(0.191)	(0.181)	(0.057)
Tongan Gr	-0.017	-0.078**	-0.054***	-0.054*	-0.056*	0.039	-0.092**	0.143	-0.111	0.119**	0.057	0.017	-0.133*	-0.144	-0.131	-0.212	-0.454**	-0.075

	(0.015)	(0.031)	(0.016)	(0.028)	(0.034)	(0.033)	(0.042)	(0.099)	(0.071)	(0.056)	(0.061)	(0.047)	(0.079)	(0.110)	(0.128)	(0.147)	(0.199)	(0.050)
Niuean Gr	-0.009	0.079**	-0.006	-0.005	-0.036	-0.098**	0.121	-0.057	0.156	0.026	-0.035	-0.039	-0.121	-0.037	0.283	-0.074	0.368*	-0.044
	(0.022)	(0.038)	(0.022)	(0.044)	(0.041)	(0.044)	(0.075)	(0.135)	(0.156)	(0.109)	(0.090)	(0.057)	(0.098)	(0.133)	(0.193)	(0.207)	(0.204)	(0.093)
Tokelauan Gr	0.023	-0.071	-0.006	-0.058	0.122	-0.307	0.046	0.802	0.479	-0.456	0.201	-0.679*	-0.143	1.428	-1.315	-0.572	-0.001	0.315
	(0.124)	(0.198)	(0.135)	(0.257)	(0.311)	(0.319)	(0.403)	(1.449)	(0.557)	(0.647)	(0.580)	(0.387)	(0.494)	(0.920)	(1.307)	(1.402)	(1.169)	(0.308)
Fijian Gr	0.063	-0.031	0.029	-0.071	-0.044	0.399**	0.248	-0.431	0.545**	-0.302	0.167	-0.032	0.179	-0.571	-0.108	-0.521	0.195	-0.245
	(0.067)	(0.112)	(0.082)	(0.131)	(0.141)	(0.170)	(0.221)	(0.677)	(0.235)	(0.313)	(0.274)	(0.184)	(0.263)	(0.456)	(0.513)	(0.808)	(0.596)	(0.239)
Other PI Gr	-0.125	0.093	0.255**	-0.137	0.193	0.597**	0.459	-0.750	-0.171	0.448	0.070	0.070	-0.613	0.115	-0.839	-0.313	-0.240	0.245
	(0.103)	(0.163)	(0.120)	(0.214)	(0.260)	(0.287)	(0.333)	(1.054)	(0.589)	(0.504)	(0.478)	(0.349)	(0.459)	(1.144)	(1.086)	(1.023)	(1.063)	(0.349)
SE Asian Gr	0.055	0.101*	-0.010	-0.067	-0.068	0.069	0.082	-0.012	0.081	-0.204	-0.130	0.028	-0.431***	-0.026	-0.127	0.009	-0.066	-0.242
	(0.034)	(0.052)	(0.034)	(0.067)	(0.079)	(0.094)	(0.104)	(0.319)	(0.174)	(0.200)	(0.128)	(0.090)	(0.141)	(0.238)	(0.265)	(0.659)	(0.432)	(0.150)
Chinese Gr	-0.011	-0.017	-0.065***	-0.042	-0.078**	-0.053*	-0.081*	-0.084	0.020	-0.111	0.013	0.105**	-0.060	-0.297***	-0.121*	-0.216	0.139	-0.077**
	(0.012)	(0.018)	(0.016)	(0.035)	(0.032)	(0.032)	(0.042)	(0.158)	(0.054)	(0.070)	(0.057)	(0.045)	(0.045)	(0.102)	(0.070)	(0.186)	(0.115)	(0.039)
Indian Gr	-0.025	-0.013	-0.015	-0.006	-0.033	-0.114**	-0.041	0.026	-0.126	0.008	-0.067	-0.009	-0.035	0.040	-0.251*	-0.344**	0.085	-0.168***
	(0.017)	(0.023)	(0.015)	(0.034)	(0.035)	(0.044)	(0.063)	(0.115)	(0.080)	(0.052)	(0.067)	(0.045)	(0.055)	(0.093)	(0.130)	(0.146)	(0.192)	(0.056)
Other Asian Gr	0.076	-0.046	-0.261**	0.068	-0.216	-0.012	0.374	-0.919	0.375	-0.660	-0.242	-0.124	-0.061	1.390***	-0.322	0.898	-0.612	0.241
	(0.073)	(0.125)	(0.112)	(0.172)	(0.190)	(0.198)	(0.232)	(1.320)	(0.338)	(0.531)	(0.301)	(0.168)	(0.234)	(0.414)	(0.748)	(0.758)	(0.844)	(0.203)
Middle Eastern Gr	-0.113	0.139	-0.107	-0.056	0.161	0.012	0.064	-0.482	0.224	0.472*	-0.206	0.186	0.280	0.445	0.559*	0.549	-0.186	0.153
	(0.104)	(0.127)	(0.121)	(0.178)	(0.129)	(0.126)	(0.283)	(0.497)	(0.372)	(0.280)	(0.292)	(0.167)	(0.242)	(0.271)	(0.327)	(0.586)	(1.000)	(0.199)
Latin American Gr	-0.201	-0.507	-0.811***	-0.161	0.589	0.112	0.264	-1.651	-0.286	-0.035	0.149	-0.966*	-1.123	-1.894	-0.344	2.460**	-2.613	-1.247*
	(0.183)	(0.311)	(0.216)	(0.403)	(0.417)	(0.480)	(0.674)	(1.954)	(0.855)	(1.059)	(0.765)	(0.569)	(0.781)	(1.793)	(1.468)	(1.023)	(2.112)	(0.708)
African Gr	0.279	-0.286	0.255	0.562	0.909**	-0.117	0.817*	1.407	0.609	-0.436	0.0207	-0.910	-0.194	2.648***	-1.055	0.403	-0.0126	-0.371
	(0.182)	(0.274)	(0.207)	(0.472)	(0.409)	(0.527)	(0.461)	(1.537)	(1.142)	(0.939)	(0.792)	(0.612)	(0.633)	(0.977)	(0.895)	(1.612)	(1.280)	(0.519)
ONFD Gr	-0.172	-0.035	-0.030	0.086	0.160	-0.323	0.552	-0.630	0.591	-1.744**	-0.551	-0.009	0.448	0.262	0.758	-0.828	0.205	0.793**
	(0.122)	(0.194)	(0.154)	(0.259)	(0.290)	(0.321)	(0.438)	(0.859)	(0.826)	(0.870)	(0.474)	(0.359)	(0.424)	(1.032)	(0.992)	(2.066)	(1.090)	(0.368)
Observations	65,800	65,800	65,800	65,800	65,600	65,500	65,500	64,200	64,700	64,800	64,900	65,700	65,400	63,800	64,900	61,700	59,000	65,500
Pseudo R-squared	0.614	0.217	0.653	0.839	0.812	0.823	0.827	0.783	0.472	0.592	0.692	0.819	0.862	0.801	0.216	0.196	0.201	0.182

Notes:

The table reports logistic regression coefficients.

We have combined the ‘Other’ and ‘not further defined’ (among those who are European, Asian or Pacific Islanders) ethnic groups into one group ‘ONFD’. Thus, our analysis includes eighteen Level 2 ethnic groups instead of twenty-one (see Table 1).

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

Clustered Standard errors in parenthesis

Tables A1, A2 and A3 are reporting results from the same regression. We have broken down the results into different tables according to blocks of explanatory variables for easy readability.

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

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

Table A4: Clustered Logistic Regression of Locational Change

Adults				Children			
Individual - level Variables	***Moved***	***Neighbourhood-level Variables***	***Moved***	***Individual - level Variables***	***Moved***	***Neighbourhood-level Variables***	***Moved***
(1)		**(2)**		**(3)**			**(4)**
Sex	-0.044*** (0.007)	**Entropy**	0.828 (1.404)	**Sex**	-0.029* (0.016)	**Entropy**	0.214 (1.420)
Age	-0.025*** (0.001)	**NZ European Gr**	0.009 (0.006)	**Age**	-0.089*** (0.006)	**NZ European Gr**	0.004 (0.007)
NZ Born	-0.045 (0.038)	**Other European Gr**	0.015 (0.016)	**NZ Born**	-0.320*** (0.046)	**Other European Gr**	0.025 (0.017)
Years at address	-0.044*** (0.002)	**NZ Māori Gr**	0.006 (0.017)	**Years at address**	-0.013*** (0.002)	**NZ Māori Gr**	0.023 (0.018)
NZ European	0.163*** (0.015)	**Samoan Gr**	0.004 (0.029)	**NZ European**	0.018 (0.046)	**Samoan Gr**	0.002 (0.031)
Other European	0.029 (0.023)	**Cook Island Māori Gr**	0.001 (0.052)	**Other European**	0.020 (0.040)	**Cook Island Māori Gr**	-0.007 (0.054)
NZ Māori	0.0615 (0.0595)	**Tongan Gr**	0.017 (0.048)	**NZ Māori**	0.155* (0.087)	**Tongan Gr**	0.014 (0.049)
Samoan	-0.418*** (0.095)	**Niuean Gr**	0.016 (0.077)	**Samoan**	-0.249** (0.101)	**Niuean Gr**	0.003 (0.081)
Cook Island Māori	-0.419*** (0.089)	**Tokelauan Gr**	0.318 (0.461)	**Cook Island Māori**	-0.228*** (0.080)	**Tokelauan Gr**	0.283 (0.487)
Tongan	-0.318*** (0.090)	**Fijian Gr**	0.007 (0.215)	**Tongan**	-0.083 (0.097)	**Fijian Gr**	0.106 (0.226)
Niuean	-0.590*** (0.106)	**Other PI Gr**	-0.110 (0.360)	**Niuean**	-0.261** (0.121)	**Other PI Gr**	-0.055 (0.383)
Tokelauan	-0.456*** (0.127)	**SE Asian Gr**	0.063 (0.126)	**Tokelauan**	-0.129 (0.261)	**SE Asian Gr**	0.033 (0.133)
Fijian	-0.097 (0.062)	**Chinese Gr**	0.015 (0.035)	**Fijian**	-0.073 (0.124)	**Chinese Gr**	0.015 (0.035)
Other PI	-0.004 (0.088)	**Indian Gr**	0.024 (0.044)	**Other PI**	-0.703*** (0.210)	**Indian Gr**	0.040 (0.044)
SE Asian	-0.050 (0.066)	**Other Asian Gr**	0.172 (0.236)	**SE Asian**	-0.224* (0.115)	**Other Asian Gr**	0.123 (0.240)
Chinese	-0.210** (0.087)	**Middle Eastern Gr**	0.243 (0.302)	**Chinese**	-0.0897 (0.108)	**Middle Eastern Gr**	0.036 (0.275)
Indian	-0.179* (0.099)	**Latin American Gr**	-0.098 (0.636)	**Indian**	-0.106 (0.105)	**Latin American Gr**	-0.110 (0.650)
Other Asian	0.051 (0.086)	**African Gr**	0.083 (0.538)	**Other Asian**	0.104 (0.198)	**African Gr**	0.180 (0.573)
Middle Eastern	0.066 (0.107)	**ONFD Gr**	0.180 (0.400)	**Middle Eastern**	-0.170 (0.185)	**ONFD Gr**	0.360 (0.427)
Latin American	0.142 (0.160)			**Latin American**	0.289 (0.338)		
African	-0.189 (0.129)			**African**	-0.456 (0.284)		
ONFD	-0.072 (0.084)			**ONFD**	0.225 (0.184)		
Observations		403,200		**Observations**		65,800	
Pseudo R-squared		0.631		**Pseudo R-squared**		0.593	

Notes:

The table reports logistic regression coefficients.

We have combined the ‘Other’ and ‘not further defined’ (among those who are European, Asian or Pacific Islanders) ethnic groups into one group ‘ONFD’. Thus, our analysis includes eighteen Level 2 ethnic groups instead of twenty-one (see Table 1).

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

Clustered Standard errors in parenthesis

Table A4, columns 1 and 2, and Table A4 columns 3 and 4 are reporting results from the same regressions. We have broken down the results into different columns according to blocks of explanatory variables for easy readability.

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