*New Zealand Population Review, 49, 119–156.*


# Towards a Dynamic Spatial Microsimulation Model for Projecting Auckland's Spatial Distribution of Ethnic Groups

MOHANA MONDAL,[*] MICHAEL P. CAMERON,[†] AND JACQUES POOT[‡]

## 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's most culturally diverse city, in which about 40 per cent of the population is foreign born. The key elements of the microsimulation model are a module that projects residential mobility within Auckland and migration between Auckland and the rest of the world, and a module that projects mobility in ethnic identity. The model is developed and calibrated using data on 1996–2001 linked populations in the 1981–2006 New Zealand Longitudinal Censuses (NZLC). We compare the microsimulation results with the actual 2006 population in each area unit. We find that in terms of indices 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 in 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.

[*] Mohana Mondal is a senior data analyst, digital engineering, at WSP in India.

[†] Michael Cameron is a professor of economics in the School of Accounting, Finance and Economics Operations at the University of Waikato | Te Whare Wānanga o Waikato. Corresponding author: mcam@waikato.ac.nz

[‡] Jacques Poot is an emeritus professor of Population Economics at the University of Waikato | Te Whare Wānanga o Waikato.

## Whakarāpopotonga

I tēnei pepa ka whakaahua mātou i te whakawhanake, tōkarikari me te whakamana i te tauira whaihanga whāiti mokowā hihiri mō te matapae i ngā taupori mātāwaka i te wāhi iti (wae horopaki) i Tāmaki Makaurau, te tāonenui he nui rawa te kanorau ahurea o Aotearoa, i whānau ai tōna 40 ōrau o te taupori i tāwāhi. Ko ngā wāhanga matua o te tauira whaihanga whāiti he kōwae e matapae ana i te panuku kainoho i roto i Tāmaki Makaurau, te hekenga i waenga i Tāmaki Makaurau me ērā atu whenua o te ao, me tētahi kōwae ka matapae i te panuku i te tuakiri mātāwaka. Kua whakawhanakehia, kua tōkarikaritia te tauira mā te whakamahi raraunga i ngā taupori honohono i te Tatauranga Wā Roa o Aotearoa (NZLC) 1981-2006, E whakataurite ana mātou i ngā otinga whaihanga whāiti ki te taupori tūturu o te 2006 i ia wae wāhi. Ko tā mātou i kite ai mō te taha ki ngā tauine o te kōmaka kainoho whānui me te kanorau mātāwaka, kua tino tata ō mātou uara matapae ki ngā uara tūturu. I te āwhata mokowā e nui ake ai te wetehiato, e pai ana te mahi a te tauira mō te taha ki te whakarato i tētahi inenga kaumingomingo taunoa whaihanga o te kanorau mātāwaka i ngā wae wāhi, engari he iti iho tana pai ki te matapae i te wehewehenga kainoho mō tēnā, mō tēnā rōpū mātāwaka.

**Ngā kupu matua:** tauira whaihanga whāiti hihiri, tuakiri mātāwaka, whakawhitinga tauwāhi, whakawhitinga mātāwaka

## Disclaimer

The results in this paper are not official statistics. They have been created for research purposes from census unit record data in the Stats NZ Datalab. The opinions, findings, recommendations and conclusions expressed in this paper are those of the authors, not Stats NZ. Access to the anonymised data used in this study was provided by Stats 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 confidentialised to protect these groups from identification and to keep their data safe. Careful consideration has been given to the privacy, security and confidentiality issues associated with using unit record census data.

The preferences of individuals regarding their residential location constitute an important topic of study because residential location of households is one of the key components of urban dynamics. The literature on residential sorting suggests that people choose where to locate based on a variety of factors (e.g., Duncan & Duncan, 1955; Schelling, 1971; Uyeki, 1964). Patterns of residential sorting have been observed to be influenced by ethnicity and race (e.g., Ho & Bedford, 2006; Johnston et al., 2011; Mondal et al., 2021b; Schelling, 1971), educational qualification (e.g., Denton & Massey, 1988; Domina, 2006; Farley, 1977), occupational status (e.g., Duncan & Duncan, 1955; Simkus, 1978), and income (e.g., Fischer, 2003). Clearly, a better understanding of urban population dynamics is needed to provide insight into what the future spatial distribution of a population might look like and to enhance thereby the efficiency and efficacy of planning for future public services and housing demands (Cameron & Poot, 2019).

Our understanding of residential sorting, and its causes and impacts, remains relatively limited (Bruch & Maré, 2006). Better understanding of changing residential sorting patterns requires 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 current and anticipated spatial features that are apparent at different spatial scales into account when deciding on their residential location. Yet most of the research on the dynamics of individual transitions and residential sorting looks either backwards in time or focuses just on the present (Rees et al., 2017).

Ethnic diversity is an important contributor to residential sorting. Schelling (1971) noted that individuals prefer to stay in close contact with people with whom they share similar preferences, which may inter alia lead to people clustering together with others of the same ethnicity. Residential sorting may also occur in terms of other characteristics such as education, income or occupation. However, in Auckland, New Zealand – the city this paper focuses on – residential sorting of the population is stronger in terms of the self-identified ethnicity of individuals than in terms of their economic characteristics (Mondal et al., 2021b). In this context, Auckland provides an important case study of residential sorting given that this city, with a population of 1.6 million (one-third of the population of New Zealand), is one of the most culturally diverse cities in the world and also the most diverse city in New Zealand (Maré & Poot, 2022; Mondal et al., 2021b).

Projections of ethnic diversity in a city require assessing the ethnic composition of the population at the neighbourhood level (O'Sullivan, 2009). This makes the task of projecting ethnic populations more difficult. The data requirements for small-area projections are high, and the methods are currently under-developed (Cameron & 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 spatial scale, namely that of census area units, and with the maximum feasible disaggregation of ethnic groups. The model is constructed with microdata from the 1981–2006 New Zealand Longitudinal Censuses (Didham et al., 2014), yielding 1996–2001 longitudinal data on ethnicity-specific populations along with their ethnic and spatial mobility. We test our model by comparing our simulated results with 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 a fine spatial scale in New Zealand. The model uses a greater level of disaggregation of ethnicity than was done in previous studies in New Zealand, but also in many other countries. This way we aim to capture better the heterogeneity that exists within the broad ethnic groups, in terms of preferences and choices (Mondal et al., 2021b). We develop and run our model in Stata, which is in itself a novel approach to dynamic spatial microsimulation modelling. The Stata statistical software is available inside the secured Stats NZ Datalab. Hence, we can run our model in the Datalab with the original microdata rather than first having to generate a sample of anonymised synthetic unit record data that can be taken out of the Datalab. Using the original microdata avoids any potential bias that might result from creating a synthetic base population. Moreover, our approach allows us to use the entire Auckland population that could be linked in the 1996 and 2001 censuses as our base population, rather than just a sample of the population.

The remainder of the paper is organised as follows. The next section reviews different types of MSMs and how they have been used in previous research, then the following two sections describe the data and the methods we employed, respectively. After that, there is a section describing the results and the testing of the MSM model, and the paper ends with a conclusion.

## Literature Review

*Microsimulation* is a methodology to model outcomes at the micro level. The outcomes can be about people (e.g., Mot, 1992), households (e.g., Rogers et al., 2014), or firms (e.g., Moeckel, 2009). Microsimulation has become increasingly popular in recent decades as ever-increasing computing power enables a growing range of applications developed by means of rich microdata (Li & O'Donoghue, 2013). Among the many applications possible, a MSM can be used to simulate and project populations and their attributes. Simulation can be interpreted here as the process by which attributes are assigned to individual units (Lomax & Smith, 2017), informed by unit record data. The base population of a MSM either can come from a survey or can be synthesised from various data sources (Zaidi & 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 & Smith, 2004; Harding, 2007), modelling lifetime earnings distributions (Holmer et al., 2014; Smith et al., 2007), 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 & Wolfson, 2000; Wolfson & Rowe, 2013).

### *Types of MSMs*

All MSMs require microdata (Wu et al., 2011), but differ in terms of the overall set-up of the model (static or dynamic), the estimation of transition 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.

*Static MSMs* usually take a cross-section of the population at a specific point in time, and measure the immediate effects of policy changes without modelling any of the specific processes that result in changes over time (Lambert et al., 1994; Spielauer, 2011). This type of MSM has been mainly used to evaluate tax-benefit systems (Pechmen & 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 in response to increasing traditional welfare and the introduction of in-work benefits in 15 countries of the European Union in 1998,[1] and 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 & Smith, 2017). For example, Favreault and Smith (2004) designed DYNASIM3 (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. This model follows 1935–1985 birth cohorts up to 2030 (Hancock et al., 1992; Holmer et al., 2014).

Dynamic MSMs can be probabilistically dynamic or implicitly dynamic. *Probabilistically dynamic MSMs* use event probabilities to project the characteristics of each unit record in the simulated database into the future. The *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, Clarke and Wiemers (2005) used a probabilistic model to project population change from 1991 until 1996 and between 1996 and 2002 at the District Electoral Division (DED) level in Ireland. Probabilistically 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, Rossiter, et al., 2005). In contrast, *implicitly dynamic* MSMs use independent small area projections and apply static simulation techniques to create small area microdata. For example, Ballas, Rossiter, et al. 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; that is, the individual's behavioural responses to the policies are not included in the model (Bourguignon & Spadaro, 2006). Hence, any behavioural responses are considered exogenous; that is, determined outside the model. Arithmetical models have been used to examine indirect taxes and tax reforms (Creedy, 1999; Sahn & Younger, 2003), to estimate incidence of public spending in health and education (Demery, 2003), and also to compare fiscal policy effects (Atkinson et al., 1988; Atkinson et al., 2002; Callan & Sutherland, 1997). 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 & Duncan, 2002). Behavioural MSMs have been used to evaluate the effects of direct tax reforms (Blundell et al., 2000; Bonin et al., 2002; van Soest & Das, 2001) as well as indirect tax reforms (Creedy, 1999; Kaplanoglou & Newbery, 2003; Liberati, 2001). The main advantages of behavioural MSMs are their 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 generalisable to the evaluation of other policies (Zucchelli et al., 2010).

Dynamic MSMs can be represented in discrete or continuous time. In the case of *discrete-time dynamic MSMs*, each individual's characteristics are simulated at fixed time intervals. These models usually include a *transition probability matrix* for the simulations (Willekens, 2006). In New Zealand, Milne et al. (2015) developed a discrete-time dynamic 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 to build their model and transition probability estimates.[2]

*Continuous-time dynamic MSMs* treat time as continuous and are, therefore, able to estimate the time at which each event occurs. In these models, individuals are assigned characteristics that can change at any time. The continuous-time dynamic MSMs use survival functions to model the length of time that an individual will remain in his/her current state, and to simulate the timing of events (Willekens, 2006). Although these models have theoretical advantages, they have higher data requirements than discrete-time MSMs (Zaidi & 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 continuous-time dynamic 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. In an *open MSM* 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*, such as DYNACAN in Canada, which generates a new unit only when a baby is born (Zaidi & 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. For example, the SVERIGE model (Rephann, 2004, Vencatasawmy et al., 1999) was the first national-level dynamic spatial MSM, and was developed from longitudinal socio-economic information on every resident in Sweden from 1985 until 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 dynamic spatial MSM. Their model was executed for 110 years (1990–2100).

Finally, MSMs differ in terms of how the base population is created. Some MSMs use census or 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 obtained from multiple alternative sources, generated for diverse purposes that are not always directly compatible. In these cases, a *synthetic population* that closely represents the actual population is created to be the base population in the MSM (Zaidi & Rake, 2001). The synthetic unit records may be generated using existing data sets and a variety of techniques like iterative proportional fitting, linear programming or complex combinatorial optimisation methods (Ballas, 2001; Ballas & Clarke, 2000; Williamson et al., 1998). 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 the Netherlands and LifePaths in Canada use a synthetic database of unit records created using the census and other data sources (Li & O'Donoghue, 2013).

### *Previous MSMs projecting ethnic population change*

Dynamic MSMs have been used previously to project the future ethnic composition of the population of several countries. For example, Demosim is a dynamic spatial MSM developed and maintained by Statistics Canada, which has been used to project the Canadian ethno-cultural 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 insight into the projected ethno-cultural make-up of the Canadian population in 2031 at different spatial scales. Taking 20 per cent 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 ethno-cultural diversity over time, both within the Canadian-born and the foreign-born populations, 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. They worked with linked microdata from the New Zealand Longitudinal Census that covers 1981 until 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 New Zealand European/Other), considering them as time-invariant (i.e., each individual's ethnicity was 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.[3] Also, changing patterns in living arrangements, such as households 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 the changes in ethnic residential segregation for four major ethnic groups (New Zealand European, Māori, Pacific and Asian). His approach 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) and macro levels (the entire Auckland urban area).[4] Several scenarios were simulated 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 of the area units in the Auckland urban area. Results also showed that the number of area units with a majority of Asian and Māori population will increase in the future in all of the territorial authorities Ardestani studied. In the Waitākere area, there would be several area units where the Pacific Peoples 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.

In a follow-up study, Ardestani et al. (2018) used a multi-scaled agent-based model to simulate the relocation of residents in the five central territorial authorities (TAs) of the Auckland urban area. The aim was to study the dynamics of residential segregation. The authors focused again on the four major ethnic groups, and found that a high-fertility and high-migration scenario leads to lesser levels of residential segregation than a low-fertility and low-migration scenario. They also found that, in the low-fertility and low-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. With respect to New Zealand, Ardestani (2013) and Ardestani et al. (2018) did not investigate the residential segregation patterns at the area unit level, and focused only on four broad ethnic groups. This overlooks the diversity *within* these ethnic groups (especially within the Asian and Pacific Peoples ethnic groups) (Mondal et al., 2021a). Additionally, these studies 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). Our model extends this earlier work, and addresses these shortcomings to some extent.

## Data

The most recent population census in New Zealand was in 2018 and recorded a usually resident population of 4.7 million. Auckland is the most ethnically diverse metropolitan region in New Zealand and accounts for about one-third of the New Zealand population (Maré & Poot, 2022; Mondal et al., 2021b). The major ethnic groups present in Auckland in the 2018 Census were European (53.5 per cent), Asian (28.2 per cent), Pacific Peoples (15.5 per cent), Māori (11.5 per cent), MELAA (2.3 per cent),[5] and Other (1.1 per cent) (Stats NZ, 2020).[6] Because of its high ethnic diversity and relatively large population, we focus on Auckland for this microsimulation research. This ensures that there are adequate sample sizes within the ethnic groups, as well as sufficient data for estimating ethnic transitions.

We use data for the Auckland region from the 1996–2001 linked populations in the 1981–2006 New Zealand Longitudinal Censuses (NZLC) (Didham et al., 2014).[7] The longitudinal census links individual records across pairs of censuses in a deterministic way. For example, an individual with age $a$ in census year $t$ who declared to have not changed address during the intercensal period is the same person as the individual of age $a$–5 in census year $t$–5 at that address. Throughout this paper, we use 'previous' to refer to data from the first census in each intercensal period and 'current' for data from the following census. The link rate for individuals from the 1996 Census to the 2001 Census was 69.5 per cent, and for the 2001 Census to the 2006 Census was 70.3 per cent (Didham et al., 2014).[8] 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 unit record data aggregated to the area unit level, using 2013 Auckland area unit boundaries.[9] In 2013, the Auckland region comprised 413 land-based area units, 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 Stats NZ's secure data laboratory, to meet the confidentiality and security rules of the Statistics Act 1975.[10]

In New Zealand, *ethnicity* captures the ethnic group(s) 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 declared. Individuals can identify with up to six ethnic groups in the census.[11] Individuals are able to choose one or more ethnicities in each census different from any they had 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 very broad and potentially mask heterogeneity in the characteristics of the ethnic groups, particularly for the Asian and the Pacific ethnic groups (Mondal et al., 2021a). Hence, we use Level 2 ethnic groups to better capture this heterogeneity. There are a non-negligible number of individuals among those who are European, Asian or Pacific Peoples who were coded as belonging to the 'Not further defined' group or the 'Other' group. We combined these two groups for each of those three ethnicities. Hence, we have 18 rather than 21 ethnic groups 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 (Level 1) | Ethnic group code (Level 2) | Ethnic Group code description (Level 2) | Ethnic group in simulation |
|---|---|---|---|---|
| 01 | European | 10 | European not further defined | 2 |
| | | 11 | New Zealand European | 1 |
| | | 12 | Other European | 2 |
| 02 | Māori | 21 | New Zealand Māori | 3 |
| 03 | Pacific Peoples | 30 | Pacific Island not further defined | 10 |
| | | 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 | 14 |
| | | 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 have been inconsistent between censuses. For instance, the ethnicity question in 2001 differed substantially from that in 1996.[12] These inconsistencies affect particularly the European ethnic groups (including New Zealand European) and the Māori ethnic group. In the 1996 data, the count for 'Other European' 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 (Stats NZ, 2017).

Second, there has also been inconsistency in the treatment of responses of 'New Zealander' to the census ethnicity question. The standard for ethnicity statistics was developed in 2005. Previously, the New Zealander response was included in the 'European' category, and was later moved to the Other ethnicity category (Statistics New Zealand, 2007a). New Zealand Europeans were the most likely group to be calling themselves New Zealander in the 2006 Census (Brown & Gray, 2009: Statistics New Zealand, 2007b). This resulted in an increase in the Other ethnicity category, and a consequent reduction in the size and proportion of people reporting as being European or New Zealand European. 'New Zealander' was included explicitly as a new category in 2006, but not in 2001. In 2001, individuals considering themselves to be a New Zealander were likely to have been counted in the New Zealand European ethnic category (Stats NZ, 2017).

Our model incorporates intercensal migration flows. This requires that we observe the location of each individual in two successive censuses. That is problematic in the case of emigration (from Auckland to overseas), and deaths, 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 the number of deaths in Auckland to each area unit according to the area unit share of total Auckland population.[13] For in-migration (from overseas or from elsewhere in New Zealand to Auckland) and births, we identified those 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 population inflow into Auckland (due to births and inward migration) and population outflow (due to deaths and outward migration), but the inflows and outflows are not split into the contributions from migration and natural change.[14]

## Methodology

In this section, we describe the construction and calibration of a dynamic spatial MSM which can be used to project the future spatial patterns of ethnic diversity in Auckland, taking both ethnic and spatial mobility into consideration. Our model is a discrete-time (runs in five-year time steps) probabilistic (uses transitional probabilities to project forward) dynamic (includes time-varying parameters) and spatial (assigns an area unit of residence to each individual) MSM. Our model is also 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 areas in New Zealand and from other countries (in-migration).

The MSM model we describe here is a *validation model*, which uses linked 1996–2001 data from the 1986–2006 NZLC to simulate and project the population in 2006, which is then validated against actual 2006 Census data. This model can then be used to develop a *projection model* that will simulate and project the population in subsequent census years. However, projecting area unit populations after 2006 is beyond the scope of the present paper. The validation model comprises two modules: (1) an ethnic transition module, and (2) a locational transition module. For each of these two 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; that is, 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. In practice, the ethnic transition module runs first in each time step, followed by the locational transition module.

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 the ethnic group *I*; 0 = otherwise), in the current census as the dependent variable. This variable represents whether 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., 2020).

**Figure 1: Theoretical framework – Ethnic transition**

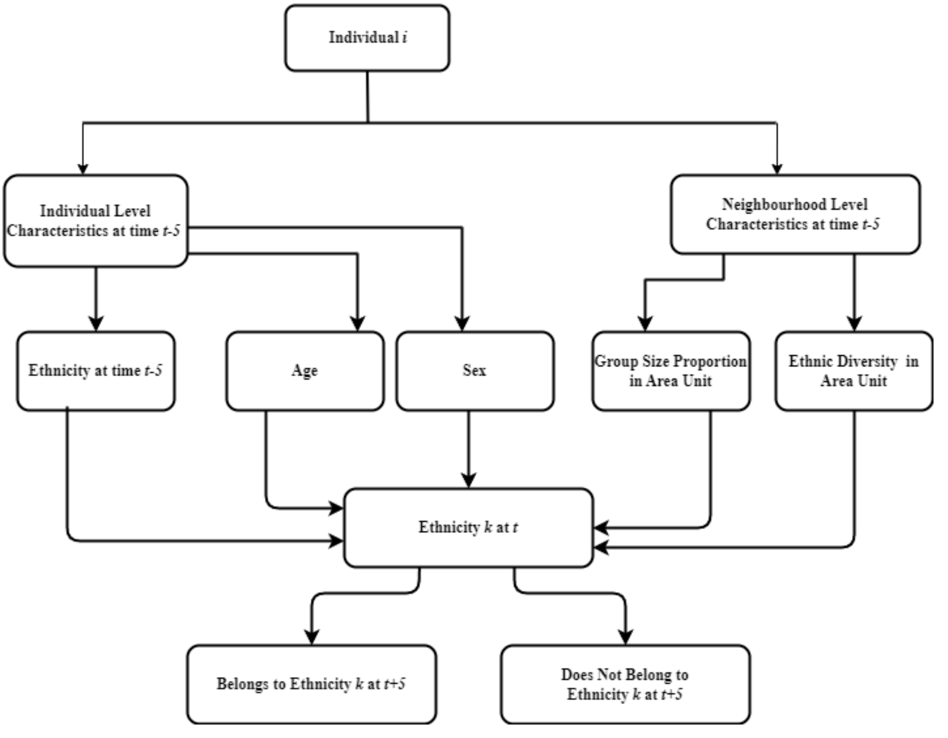

Figure 2: Theoretical framework – Location transition

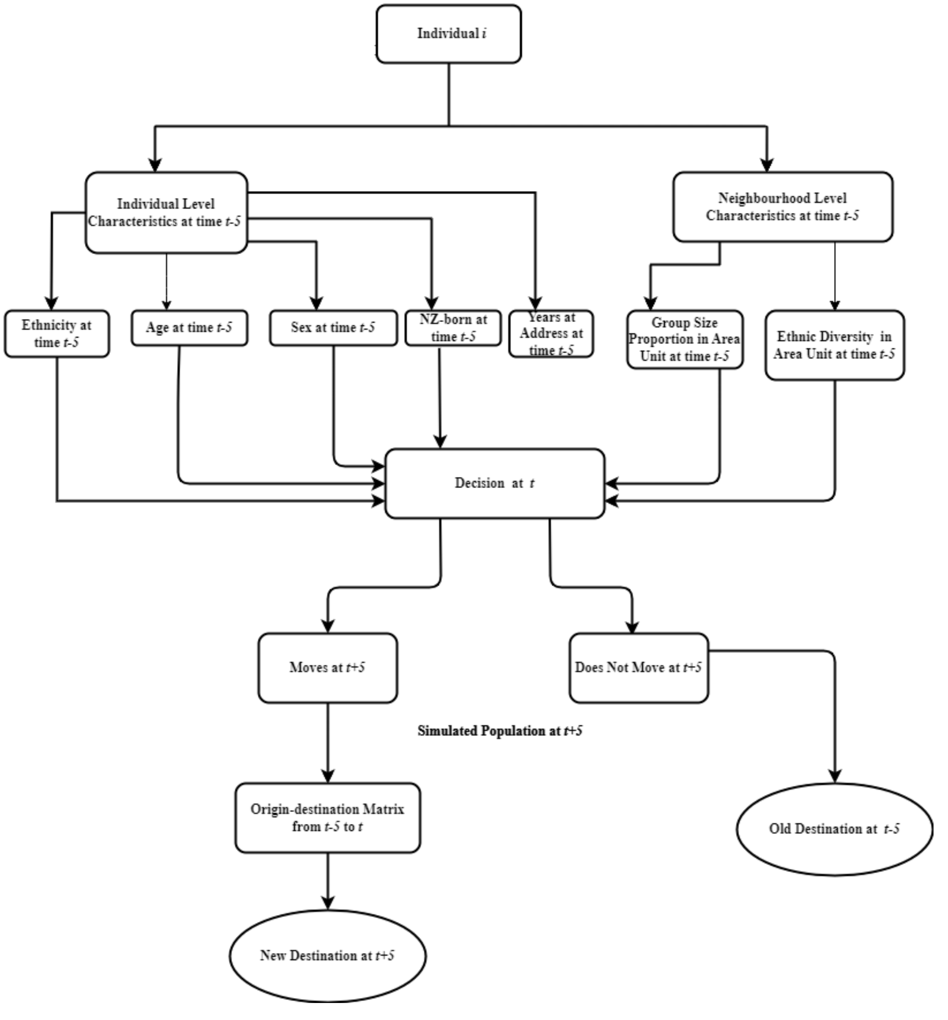

**Table 2: Variables used in the analysis**

| **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, New Zealand-born |
| | | Neighbourhood | Ethnic diversity in area unit, Ethnic group size proportions in area unit |
| Location transition | Moved[20] (1 = moved; 0 = otherwise) | Individual | Ethnicity, Age, Sex, New Zealand-born, years at address |
| | | Neighbourhood | Ethnic diversity in area unit, Ethnic group size proportions in area unit |

Notes: 1. These logit models are estimated separately for the population aged 0–17 and the population aged 18 and over.

2. We created the binary variable 'moved' (1 = if individual changed area unit during the intercensal period; 0 = otherwise) from the census data on location of usual residence in the current census and the variable 'address five years ago' for the same individual.

An individual's ethnicity in our model is an 18 x 1 row vector of binary variables, with one binary variable for each of the 18 ethnic groups *i*. Our approach allows us to include multiple ethnic affiliations for individuals without requiring an order of priority for the determination of the ethnic choices; that is, 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. 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 any of the possible ethnicities in the projected year.

The individual-level determinants of ethnicity in 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. 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 in the logistic regressions were observed at the start of each intercensal 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. Specifically, we first use logistic regression equations (with separate coefficients for adults and 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 in the current year.

In the second stage, the people who changed their location are then distributed over possible destinations using a distribution function that is solely dependent on the destination but not on the origin. In this step, movers are allocated to destination area units based on a column-standardised origin-destination matrix (with a zero diagonal) calculated using the intra-urban relocation data from the actual 1996–2001 linked census. 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.[15] The destination for each migrant is determined again using a uniformly distributed random variable, with the appropriate column of the origin-destination matrix used as a look-up table to determine the selected destination probabilistically. Those individuals where 'outside Auckland' (out-migration or death) is selected as the destination are removed from the data set.

As the decision to move is affected by duration of stay (Poot, 1987), we include the number of years the resident has lived in the origin area unit as an explanatory variable in the locational transitional equations along with all variables included in the ethnic transition equations.

## *Simulation evaluation*

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 with 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.

### Measures of residential sorting

There are many different measures that can be used as indicators of residential sorting; see, for example, Massey and Denton (1988), Nijkamp and Poot (2015), and 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 (Mondal et al., 2021a; Reardon & Firebaugh, 2002). The measures used in our analysis are detailed in Table 3. In order to observe the extent to which ethnic groups are over- or underrepresented 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, $I_a$, which varies between 0 and 1. The value of the *diversity evenness index* is 0 (i.e., $E_a = 0$) when only one of the groups is present in area unit $a$, and is 1 (i.e., $E_a = 1$) when all groups are equally represented in area unit *a* (Nijkamp & Poot, 2015). We also use the *entropy index of spatial sorting of group g* ($EIS_g$), which measures the area-population weighted average of 1 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 0 (when the group is distributed proportionally to the total population in all area units) and 1 (when all areas in which group *g* is represented contain no other group). We also calculate an overall measure of residential sorting *(H*),* by taking the group-population weighted average of the $EIS_g$ values. This is an alternative way of calculating the *Theil's Multi-group Segregation Index H* (Theil, 1972; Theil & 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.[16] The normalised diversity index ranges from 0 (when only one ethnic group is present in the area unit) to 1 (when all ethnic groups are equally represented in area unit) (Nijkamp & Poot, 2015).

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

**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) | $I_a = -\frac{\sum_{g=1}^{G}\frac{P_{ga}}{P_a}\ln\frac{P_{ga}}{P_a}}{\ln(G)}$ |
| *Normalised entropy diversity* (city) | $I^{*} = -\frac{\sum_{g=1}^{G}\frac{P_g}{P}ln\frac{P_g}{P}}{ln(G)}$ |
| *Entropy index of segregation* (group) | $EIS_g = \sum_{a=1}^{A}\frac{P_a}{P}\left(1-\frac{E_{ga}}{\bar{E}_g}\right)$ |
| where: | $E_{ga} = -\frac{P_{ga}}{P_a}\ln\left(\frac{P_{ga}}{P_a}\right) - (1-\frac{P_{ga}}{P_a})\,ln\left(1-\frac{P_{ga}}{P_a}\right)$<br>$\bar{E}_g = -\frac{P_g}{P}ln\left(\frac{P_g}{P}\right) - (1-\frac{P_g}{P})\,ln\left(1-\frac{P_g}{P}\right)$ |
| *Theil's multi-group spatial sorting index (city)* | $H^{*} = \sum_{g=1}^{G}\frac{P_g}{P}EIS_g$ |

Notes: 1. $P_{ga}$ refers to the population of group $g$ ( = 1, 2, … $G$) in area $a$ ( = 1, 2, ... $A$).
$P_a$ is the total number of people in area unit $a$.
$P_g$ is the number of members of group $g$ in Auckland and $P$ is the total population of Auckland.

2. Comparing group $g$ with all other groups combined, we denote the entropy of area $a$ as ($E_{ga}$) and of the whole Auckland city as $\bar{E}_g$.

3. The calculation of EIS requires that we define 0*ln(1/0) = $\lim_{q\to 0}[-q\ln(q)] = 0$ 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).

from less than 9 to over 3000. WMAPE at projected year $t$ is defined as:

$$WMAPE_{t-5,t} = \sum_{g}\left(\left|PE_{t-5,t}^{g}\right|\frac{P_{gt}}{P_t}\right)$$

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 & Swanson, 1999). Thus, we also report the median absolute percentage error (MedAPE$_t$) and the median algebraic percentage error (MedALPE$_t$), neither of which is 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, 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.

## *Calibration process*

After performing the initial stages of model coding and running, we calibrated the model so that the simulated 2006 population using the 1996–2001 linked data in the NZLC would be as close as possible to the actual 2006 population. We expect that if the simulated proportion of people changing their location, the proportion of people in each ethnic group, and the proportion of each ethnic group changing their location are close to the actual proportions, then the model should be able to replicate the actual levels of ethnic diversity and residential sorting in the Auckland population in 2006. 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 of staying at the current location, thereby ensuring that the model would decrease the number of 'movers'. The model then uses this calibrated random variable to calculate the predicted probabilities to determine whether a person is a mover.

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 simulation generated too many members in 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 simulation generated 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. 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.

## Results

The ultimate aim of the dynamic spatial MSM model is to be a *projection model* that will project the population forward with errors that remain small enough for the results to be useful for informing local public policy and urban management. The outcome depends strongly on the extent to which we can accurately model transitions. To obtain the predicted probabilities for both ethnic transition and location transition, we ran logistic regression equations with clustered standard errors.[17]

There are too many coefficients to discuss the logistic regression results in detail. However, there are some general patterns that provide insight into the determinants of location and ethnicity transitions. Most generally, the coefficients often differ between adults and children (those aged less than 18 years).[18] The logistic regression of intra-urban mobility shows that New Zealand Europeans are more mobile than average while those with Pacific Island ethnicity are less mobile. As expected, residential mobility declines with age and with duration of residence. Females are less mobile. Ethnic diversity of area units and the various ethnic-group shares do not appear to influence the rate of intra-urban mobility. However, New Zealand-born children and adolescents are less mobile than others aged less than 18 years.

With respect to ethnic mobility, there is, as expected, a lot of persistence: the most important predictor of ethnicity at time $t$ is ethnicity at time $t-5$. There are also some interesting correlations between ethnic groups. For example, having identified as Other European at the previous census has a positive effect on identifying as a New Zealand European in the current census. Similarly, having identified as Asian or from the Pacific in the previous census generally reduces the likelihood of identifying as Other European ethnicity in the current census. Ethnic mobility is lower at older ages and among the New Zealand-born; that is, the non-immigrants. High ethnic diversity of an area unit (i.e., a relatively large value of the entropy diversity index) leads to a greater likelihood of identifying as Other European, Samoan or Middle Eastern ethnicity. A large 'own group' share of the area unit population, however, does not always imply a stronger identification with that group – in fact the opposite is sometimes true. For example, in areas where the share of New Zealand European or of Other European is large, the likelihood of declaring these respective ethnicities is lower.

We validated the ability of the current model to replicate known 2006 Census outcomes. Table 4 shows that 21 per cent of the people who were in Auckland in 2001 and 2006 changed at least one of their identified ethnicities during the intercensal period, and the proportion is very similar for the simulated 2006 Auckland population, at 22 per cent. Likewise, the percentage of people reporting moving from one area unit in 2001 to a different area unit in 2006 was 40 per cent in the 2006 Census and the simulated percentage is 42 per cent; again, very similar. The difference in the percentage of people moving out of Auckland between the actual and the simulated data is 3 percentage points, being 9 per cent and 6 per cent, respectively.

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 close to the actual value, the difference being −0.008 (or 9.7 per cent). 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 4: Comparison between simulated data and the actual Census 2006 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 based on the simulated 2006 Census data and the actual 2006 Census data.

**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 $EIS_g$ and the normalised entropy diversity measure for area units $I_a$. The WMAPE is smaller than the MedAPE for the simulated spatial sorting/segregation of the ethnic groups (19.34 per cent and 28.53 per cent, respectively). The fact that the MedALPE has the same absolute value as the MedAPE indicates that the simulation underestimates group segregation for all groups.

The negative MedALPE value (−28.53 per cent) reflects, therefore, that there is downward bias in the simulated values of the entropy index of segregation measure, potentially resulting from the fact that not all determinants of ethnic mobility have been observed. The inconsistencies in the ethnic categorisations in the 1996 and 2001 census data mentioned earlier, 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 measures of overall ethnic residential sorting in

**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 and *I* refers to normalised entropy diversity (area unit).

Auckland are very similar (Table 5), the model does not perform as well when we simulate the ethnic residential sorting for individual ethnic groups.

With respect to the diversity measure, the WMAPE is larger than the MedAPE, which is in turn larger than the MedALPE (4.07 per cent, 3.54 per cent and 1.68 per cent, respectively). It is clear that the simulation performs better in projecting the diversity of areas than the spatial sorting of ethnic groups.

## Conclusion

The main aim of this paper is 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 1986–2006 NZLC were used to simulate the spatial distribution of the Auckland population by ethnicity in 2006. The simulated results were then compared with the actual 2006 Census data.

We have demonstrated that census data can be used to inform, calibrate and validate our model. Our simulation is generally capable of reproducing the dynamics of residential sorting in Auckland without requring detailed information on all the elements of an individual's residential decision-making process. Projection errors vary with population size of a region (Tayman et al., 1998; Smith & Shahidullah 1995). Smith and Shahidullah worked on projections of total population for all census tracts in three counties in Florida (Dade, Duval and Pinellas) and found that error measure values decline with increase in population size. Their reported mean absolute percentage errors (MAPEs) ranged from 17.3 per cent to 27.6 per cent. Tayman et al., in their work on census tracts projections in San Diego County, reported that in the census tracts with population size between 1000 and 1500, 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 our work is around 1500 on average, the results show that our model projects the spatial distribution of ethnicities in Auckland with a reasonable level of error.

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 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 simulation is the whole Auckland population in the 2001 Census, whether linked in the 1996–2001 NZLC data or not. Thus, any extent to which unlinked and linked people differ in ways that are correlated with the transitions we estimate will generate bias in the results. However, some of this bias will be attenuated through the process of calibration.

Second, due to 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 subgroups 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.

Third, an individual's location decision and ethnic choices are dependent on a variety of factors in addition to 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 are available in the census, the same data for children transitioning to adulthood are 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.[19] 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 including an additional educational transition module to the model alters the results.

Fourth, ethnic identity of the parents is important for the evolution of ethnic identity of adolescents (Mondal et al., 2020). However, the NZLC does not have this data for all children, only for children living at home with their parents (who may not be their biological parents). Moreover, the linkage rate between censuses is lower for children than for adults. Given these challenges, we chose to infer parental ethnicity using the ethnicity of all adults, rather than having differential bias between children who could and could not be linked with their parents (which may in turn be correlated with parental ethnicity).

In spite of these limitations, this paper has described the development of a modelling approach to project urban ethnic diversity at a fine spatial scale and relatively narrowly defined ethnic groups. Our model was developed using Stata, which extends the number of resources previously used to build and run microsimulation models. Our future focus will be to use this calibrated model, 2013–2018 NZLC data and the 2023 Census data when they become available to project the future ethnic spatial distribution in Auckland forward to 2038.

## Notes

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

2 These studies are the Christchurch Health and Development Study, the Dunedin Multidisciplinary Health and Development Study, and the Pacific Islands Families Study.

3 See also Mondal et al. (2021b), who show similar past trends for Auckland.

4 The territorial authorities considered were Auckland City, Manukau, North Shore, Waitakere and Papakura.

5 Middle Eastern/Latin American/African.

6 Percentages do not sum to 100 per cent, as people can report more than one ethnicity.

7 Data from the 2018 Census have not yet been integrated into the NZLC data set. Work has been undertaken to link data from the 2013 Census

to the 2006 Census (Kang, 2017), but these data were unavailable at the time of writing.

8 The link rate for the 2006 Census and 2013 Census is 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 (Didham et al., 2014).

9 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.

10 As stated in the Disclaimer at the start of this paper.

11 Individuals could identify with up to three ethnic groups until the 1996 Census, then up to to six in later censuses.

12 In the 1996 Census, the ethnicity question had a different format compared with that used in the 1991 Census and 2001 Census. In the 1996 Census, 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 Census and 2001 Census. Moreover, the first two answer boxes appeared in a different order in the 1996 Census from that in the 1991 Census and 2001 Census: in the 1996 Census, New Zealand Māori was listed first and New Zealand European or Pākehā was listed second. Another difference is that the ethnicity questions in the 1991 Census and 2001 Census only used the words New Zealand European whereas the 1996 Census used 'New Zealand European or Pākehā' (Pākehā is the Māori word referring to a person of European descent). Furthermore, the 2001 Census ethnicity question used the word Māori rather than New Zealand Māori (Stats, 2017).

13 Total emigration was calculated as a residual of 1996–2001 Auckland population change after accounting for recorded births, deaths and internal migration.

14 Intercensal births can of course only affect the age group 0–4 years in the current census.

15 We use a randomly selected ethnicity, as there is no empirical basis for selecting a particular ethnic-specific origin-destination matrix for each individual.

16 Despite the entropy-based diversity and sorting measures requiring us to take the natural logarithm of population shares when certain groups may be absent from certain areas, this does not cause a computational problem because $-\frac{P_{ga}}{P_a}\ln\frac{P_{ga}}{P_a} = 0$ when $P_{ga} = 0$, given that 0*ln(1/0) = $\lim_{q \to 0}[-q(\ln(q)] = 0$. See also the notes at the bottom of Table 3.

17 Tables of the logistic regression results are available from the corresponding author on request.

18 However, no formal statistical tests of equality of coefficients were conducted.

19 Further details are available from the corresponding author on request.

## References


Ardestani, B. M. (2013). *Using a hybrid model for investigating residential segregation: An empirical and simulation-based study* (Doctoral thesis, University of Auckland, New Zealand). ResearchSpace. 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. http://dx.doi.org/10.1016/j.compenvurbsys.2017.11.002

Atkinson A. B., 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. http://dx.doi.org/10.1016/0014-2921(88)90178-X

Atkinson, A. B., 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. https://doi.org/10.1111/1468-0335.00281

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. P. (2000). GIS and microsimulation for local labour market policy analysis. *Computers, Environment and Urban Systems*, *24*, 305–330. http://dx.doi.org/10.1016/S0198-9715(99)00051-4

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

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

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. http://dx.doi.org/10.1111/j.1475-5890.2000.tb00581.x

Bonin, H., Kempe, W., & Schneider, H. (2002). *Household labour supply effects of low-wages subsidies in Germany* (IZA Discussion Paper 637). IZA – Institute for the Study of Labor. https://www.iza.org/publications/dp/637/household-labor-supply-effects-of-low-wage-subsidies-in-germany

Bourguignon, F., & Spadaro, A. (2006). Microsimulation as a tool for evaluating redistribution policies. *Journal of Economic Inequality*, *4*, 77–106. http://dx.doi.org/10.1007/s10888-005-9012-6

Brown, P., & A. Gray (2009). Inter-ethnic mobility between the 2001 and 2006 Censuses: The statistical impact of the New Zealander response. In *Final Report of a Review of the Official Ethnicity Standard 2009* (pp 27–36). Statistics New Zealand.

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

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

Callan, T., & Sutherland, H. (1997). The impact of comparable policies in European countries: Microsimulation approaches. *European Economic Review*, *41*(3–5), 327–333. http://dx.doi.org/10.1016/S0014-2921(97)00006-8

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), 195–216. https://www.anzrsai.org/assets/Uploads/PublicationChapter/AJRS-23.2-pages-195-to-216.pdf

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. https://population.org.nz/wp-content/uploads/2019/12/NZPR-45_whole-doc-final.pdf

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. https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/journals-and-magazines/social-policy-journal/spj36/36-self-identified-ethnicity-change-over-time.html

Creedy, J. (1999). *Modelling indirect taxes and tax reform*. Edward Elgar.

Creedy, J., & Duncan, A. (2002). Behavioural microsimulation with labour supply responses. *Journal of Economic Surveys*, *16*, (1),1–39. https://doi.org/10.1111/1467-6419.00158

Davis, P., & Lay-Yee, R. (2019). *Simulating societal change: Counterfactual modelling for social and policy inquiry*. Springer.

Demery, L. (2003). Analysing the incidence of public spending. In F. Bourguignon & 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.

Denton, N. A., & Massey, D. S. (1988). Residential segregation of blacks, Hispanics and Asians by socioeconomic status and generation. *Social Science Quarterly, 69*(4), 797–817. https://doi.org/10.2307/1392486

Didham, R., Nissen, K., & Dobson, W. (2014). *Linking censuses: New Zealand longitudinal census 1981–2006.* Stats NZ. https://www.stats.govt.nz/methods/linking-censuses-new-zealand-longitudinal-census-19812006/#:~:text=This%20report%20presents%20the%20results,of%20population%20change%20across%20time

Didham, R. (2016). Ethnic mobility in the New Zealand census, 1981–2013: A preliminary look. *New Zealand Population Review*, 42, 27–42. https://population.org.nz/wp-content/uploads/2017/06/Vol-42-Full-document_Final.pdf

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

Duncan, O. D., & Duncan, B. (1955). Residential distribution and occupational stratification. *American Journal of Sociology, 60*, 93–503. https://doi.org/10.1086/221609

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. https://doi.org/10.1007/s10198-015-0714-9

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. https://doi.org/10.2307/2060592

Favreault, M., & Smith, K.E. (2004). *A primer on the dynamic simulation of income model (DYNASIM3).* The Urban Institute.

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

Goldman, D. P., Zheng, Y., Girosi, F., Michaud, P. C., Olshansky, S. J., Cutler, D., & Rowe, J. W. (2009). The benefits of risk factor prevention in Americans aged 51 years and older. *American Journal of Public Health,* 99(11), 2096–2101. https://doi.org/10.2105%2FAJPH.2009.172627

Hancock, R., Mallender, J., & Pudney, S. (1992). Constructing a computer model for simulating the future distribution of pensioners' incomes for Great Britain. In R. Hancock & H. Sutherland (Eds.), *Microsimulation Models for Public Policy Analysis: New Frontiers* (pp. 33–66). 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.

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

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. Available from: https://www.researchgate.net/publication/253561368_The_SVERIGE_Spatial_Microsimulation_Model

Holmer, M., Janney, A., & Cohen, B. (2014). *PENSIM overview*. U.S. Department of Labor, Office of Policy Research. Available from: https://www.researchgate.net/publication/265922472_Pensim_Overview

Iceland, J., Weinberg, D. H., & Steinmetz, E. (2002). *Racial and ethnic residential segregation in the United States 1980–2000, Appendix B*. US Census Bureau. https://www.census.gov/content/dam/Census/library/publications/2002/dec/censr-3.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. http://dx.doi.org/10.1111/j.1468-0297.2007.02000.x

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

Kang, I. (2017). New Zealand longitudinal census development. *International Journal of Population Data Science*, *1*(1) (Proceedings of the IPDLN Conference (August 2016). https://doi.org/10.23889/ijpds.v1i1.252

Kaplanoglou, G., & Newbery, D. M. (2003). Indirect taxation in Greece: Evaluation and possible reform. *International Tax and Public Finance*, *10*, 511–533. https://doi.org/10.1023/A:1026102606063

King, A., Baekgaard, H., & Robinson, M. (1999). *Dynamod2: An overview* (Technical paper no. 19). National Centre for Social and Economic Modelling (NATSEM), University of Canberra.

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

Li., J., & O'Donoghue, C. (2013). A survey of dynamic microsimulation models: Uses, model structure and methodology. *International Journal of Microsimulation*, *6*(2), 3–55. http://dx.doi.org/10.34196/ijm.00082

Liberati, P. (2001). The distributional effects of indirect tax changes in Italy. *International Tax and Public Finance*, *8*(1), 27–51. http://dx.doi.org/10.1023/A:1008789432330

Lomax, N., & Smith, A. (2017). Microsimulation for demography. *Australian Population Studies*, *1*(1), 73–85. http://dx.doi.org/10.37970/aps.v1i1.14

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

Maré, D. C., & Poot, J. (2022). *Accounting for social difference when measuring cultural diversity* (Motu Working Paper 22-04). Motu Economic and Public Policy Research. https://motu-www.motu.org.nz/wpapers/22_04.pdf

Massey, D. S., & Denton, N. A. (1988). The dimensions of racial segregation. *Social Forces, 67*, 281–315. https://doi.org/10.2307/2579183

Milne, B., Lay Yee, R., McLay, J. M., Pearson, J., Von Randow, M., & Davis, P. (2015). Modelling the early life-course (MELC): A microsimulation model of child development in New Zealand. *International Journal of Microsimulation, 8*(2), 28–60. https://microsimulation.pub/articles/00116

Moeckel, R. (2009). Simulation of firms as a planning support system to limit urban sprawl of jobs. *Environment and Planning B: Planning and Design*, *36*, 883–905. http://dx.doi.org/10.1068/b34110

Mondal, M., Cameron, M. P., & Poot, J. (2020). *Determinants of ethnic identity among adolescents: Evidence from New Zealand* (Working Paper in Economics No. 5/20). University of Waikato. https://www.waikato.ac.nz/__data/assets/pdf_file/0006/672144/Determinants-of-ethnic-identity-among-adolescents-Evidence-from-New-Zealand.pdf

———— (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* (pp. 113–136). Springer Nature.

———— (2021b). Cultural and economic residential sorting of Auckland's population, 1991–2013: An entropy approach. *Journal of Geographical Systems, 23*(2), 291–330. https://link.springer.com/article/10.1007/s10109-020-00327-1

Mot, E. S. (1992). *Survey of microsimulation models:* Inventory and recommendations. Social Security Research Committee. VUGA.

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

O'Sullivan, D. (2009). Changing neighbourhoods—neighbourhoods changing: A framework for spatially explicit agent-based models of social systems. *Sociological Methods & Research*, *37*(4), 498–530. http://dx.doi.org/10.1177/0049124109334793

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). https://www.iser.essex.ac.uk/wp-content/uploads/files/working-papers/euromod/em8-09.pdf

Pechman, J., & Okner, B. (1974). *Who bears the tax burden?* 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. http://dx.doi.org/10.1111/j.1745-7939.1987.tb01205.x

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

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. http://dx.doi.org/10.1016/j.ssresearch.2008.10.002

Rees, P. H., Wohland, P., Norman, P., Lomax, N., & Clark, S. D. (2017). Population projections by ethnicity: Challenges and solutions for the United Kingdom. In D. A. Swanson (Ed.), *The frontiers of applied demography* (Vol. 9, pp. 383–408). Springer.

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

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. http://dx.doi.org/10.34196/ijm.00102

Rowe, G., & Wolfson, M., (2000, June 8–9). *Public pensions – Canadian analysis based on the Lifepaths generational accounting framework* [Paper presentation]. 6th Nordic Seminar on Microsimulation Models, Copenhagen, Denmark. Revised paper available from: https://www.researchgate.net/figure/fig1_228774727

Sahn, D., & Younger, S. (2003). Estimating the incidence of indirect taxes in developing countries. In F. Bourguignon & L. A. Pereira da Silva (Eds.), *The impact of economic policies on poverty and income distribution: Evaluation techniques and tools* (pp. 27–40). World Bank & Oxford University Press. http://hdl.handle.net/10986/15090

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

Siegel, J. S. (2002). *Applied demography: Applications to business, government, law and public policy*. Academic Press.

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

Smith, S. K., & Shahidullah, M. (1995). An evaluation of population projection errors for census tracts. *Journal of the American Statistical Association, 90*(429), 64–71. http://dx.doi.org/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

Spielauer, M. (2011). What is social science microsimulation? *Social Science Computer Review*, *29*(1), 9–20. http://dx.doi.org/10.1177/0894439310370085

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

Statistics New Zealand. (2007a). *Profile of New Zealander responses, ethnicity question: 2006 Census*. Retrieved 10 October 2019 from https://statsnz.contentdm.oclc.org/digital/api/collection/p20045coll20/id/4/download

——— (2007b). *Survey of dynamics and motivations for migration in New Zealand – information releases*. https://datainfoplus.stats.govt.nz/Item/nz.govt.stats/843abe2e-aa44-496b-a4cb-da54d32e0ba3/28

——— (2013). *2013 Census definitions and forms.* https://www.stats.govt.nz/methods/2013-census-definitions-and-forms

——— (2015). *IDI data dictionary: 2013 Census data*. https://datainfoplus.stats.govt.nz/Item/nz.govt.stats/571d9fc4-ffb2-4422-b10d-ee5f05eec2ce/28

Stats NZ. (2017). *2001 Census of Population and Dwellings: Change in ethnicity question.* Retrieved 12 February 2020 from http://www.stats.govt.nz/Census/2001-census-data/change-in-ethnicity-question.aspx

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

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. http://dx.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. http://dx.doi.org/10.1023/A:1006166418051

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

Theil, H., & Finezza, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts. *Journal of Mathematical Sociology*, *1*, 187–194. http://dx.doi.org/10.1080/0022250X.1971.9989795

Uyeki, E. S. (1964). Residential distribution and stratification, 1950–1960. *American Journal of Sociology*, 69, 491–498. http://dx.doi.org/10.1086/223652

van Soest, A., & Das, M. (2001). Family labour supply and proposed tax reforms in the Netherlands*. De Economist*, *149*, 191–218. http://dx.doi.org/10.1023/A:1017531304601

Vencatasawmy, C. P., Holme, K., Rephann, T., Esko, J., Swan, N., Öhman, M., Åström, M., Alfredsson,E., & Siikavaara, J. (1999, September 3–7). *Building a spatial microsimulation model* [Paper presentation]. 11th European Colloquium on Quantitative and Theoretical Geography, Durham, England. https://www.norrbottenskommuner.se/media/1552/building-a-spatial-microsimulation-model.pdf

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.

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.org/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. https://doi.org/10.1068/a300785

Wilson, T. (2012). Forecast accuracy and uncertainty of Australian Bureau of Statistics state and territory population projections. *International Journal of Population Research, 2012*, 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*(2), 253–275. https://www.anzrsai.org/assets/Uploads/PublicationChapter/Wilson-2015-2.pdf

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. https://doi.org/10.1016/S0140-6736(13)61402-6

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

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. http://dx.doi.org/10.1177/0894439310370113

Zaidi, A., & Rake, K. (2001). *Dynamic microsimulation models: A review and some lessons for SAGE* (Simulating Social Policy in an Ageing Society (SAGE), Discussion Paper. 2). London School of Economics. Available from: https://www.researchgate.net/publication/265280929_Dynamic_Microsimulation_Models_A_Review_and_Some_Lessons_for_SAGE

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. https://www.york.ac.uk/edia/economics/documents/herc/wp/10_03.pdf