A Longitudinal Typology of Neighbourhood-level Social Fragmentation: A Finite Mixture Model Approach

Peter Lekkas ^a Natasha J Howard ^b Ivana Stankov ^c Mark Daniel ^d Catherine Paquet ^a

Abstract

Neighbourhoods are social enclaves. And, from an epidemiological vantage there is substantive research examining how social traits of neighbourhoods affect health. However, this research has often focused on the effects of social deprivation. Less attention has been given to social fragmentation (SF), a construct aligned with the notions of lesser: social cohesion, social capital, collective functioning, and social isolation. Concurrently, there has been limited research that has described the spatial and temporal patterning of neighbourhood-level social traits. With a focus on SF the main aims of this paper were to model and describe the time-varying and spatial nature of SF.

Conceptually, this research was informed by 'thinking in time' and by the 'lifecourse-of-place' perspective. While, from an analytical perspective, a longitudinal (3-time points over 10-years) neighbourhood database was created for the metropolitan region of Adelaide, Australia. Latent Transition Analysis was then used to model the developmental profile of SF where neighbourhoods were proxied by 'suburbs', and the measurement model for SF was formed of 9-conceptually related census-based indicators. A four-class, nominal-level latent status model of SF was identified: class-A=low SF; class-B=mixed-level SF/inner urban; class-C=mixed-level SF/peri-urban; and class-D=high SF. Class-A and -D neighbourhoods were the most prevalent at all time points. And, while certain neighbourhoods were inferred to have changed their SF class across time, most neighbourhoods were characterised by intransience.

Key words

Latent variable model, Longitudinal, Neighbourhood, Social environment, Urban change

Corresponding author, and author institutional affiliations

Peter Lekkas ^a lekpy001@mymail.unisa.edu.au ****corresponding author****
Natasha J Howard ^b natasha.howard@sahmri.com
Ivana Stankov ^c is379@drexel.edu
Mark Daniel ^d mark.daniel@unisa.edu.au
Catherine Paquet ^a catherine.paquet@unisa.edu.au

^a Australian Centre for Precision Health and School of Health Sciences, University of South Australia

^b formerly at Sansom Institute Health Research Operations, Division of Health Sciences, University of South Australia (current affiliation is at the South Australian Health & Medical Research Institute)

^c Urban Health Collaborative, Dornsife School of Public Health, Drexel University, U.S. of America

^d School of Health Sciences University of South Australia, Australia

1.1 Introduction and Background

Neighbourhood change has long motivated enquiry. Indeed, many contemporary studies into neighbourhood change resonate with research from the Chicago School of Urban Sociology in the early-to-mid twentieth century (Harris and Feng 2016). Prominent within the Chicago School was a perspective that cities were social ecologies and neighbourhoods natural areas that evolved and regressed through social mobility processes such as invasion and succession (Lutters 1996, Schwirian 1983). Since, the study of neighbourhood change has been advanced via a range of theories, these variously grounded in subjects that include economics, humanism, individualism, complexity, justice and political economy (Clark 2008, Hirsch et al. 2017, Joseph 2008, Meen 2013, Schwirian 1983).

In spite of interest in neighbourhood change, much of the applied research has arguably been “unidimensional”, focused on singular neighbourhood-level traits, for example population density, income, poverty, housing, race, ethnicity, crime or violence (Delmelle 2015, p.1). Treating neighbourhoods as unidimensional can overlook the “bundle of spatially based attributes” that come together to define the character of a neighborhood (Galster 2001, p.2112); it can also mask within- and across-neighbourhood variability on the basis of co-varying neighbourhood-level characteristics.

However, there is a growing body of research that addresses neighbourhood change from a multi-dimensional vantage. Much of this research applies a range of methods to capture the multidimensional character of neighbourhoods *in* space and time as well as *across* space and time through the construction of neighbourhood typologies formulated from arrays of indicators. Examples of applied methods for the multivariate study of neighbourhood change include cluster analysis (Mikelbank 2011, Morenoff and Tienda 1997), self-organising maps (Delmelle et al. 2013, Ling and Delmelle 2016), latent class models (Richardson et al. 2014, Weden et al. 2011), Markov models (Delmelle and Thill 2014, Delmelle et al. 2016), transition matrices (Solari 2012) and, latent class growth analysis (Apparicio et al. 2015, Séguin et al 2015); applied either alone or in step-wise combination with tools such as discriminant analysis (Wei and Knox 2014), principal components analysis (Owens 2012, Salvati et al. 2018), and sequential pattern mining algorithmic techniques (Delmelle 2016).

Although researchers are attentive to capturing neighbourhood change from a multidimensional vantage, their focus has, in general, remained on expressing the longitudinal multivariate socioeconomic and or sociodemographic character of neighbourhoods. Moreover, many multivariate studies have examined neighbourhood change either through comparisons of cross-sectional neighbourhood profiles modelled repeatedly across time or via the measurement and linkage of neighbourhoods across only two time points (Delmelle 2016, Weden et al 2011, Wei and Knox 2014).

1.2 Aims

This study aims to model the neighbourhood-level evolution of social fragmentation from 2001 to 2011, across three time points, within the urban metropolitan context of the city of Adelaide, Australia. Research questions examined are: What is the typological nature of neighbourhood-level social fragmentation? And, to what extent do neighbourhoods transition into different states of social fragmentation (categories) over time?

1.3 A 'Neighbourhood-Centred' Latent-Variable Approach to Change

Neighbourhoods are arguably multifaceted, if not complex constructs that may not easily reduce to a single indicator or an observable metric (Lekkas et al. 2017a, 2017b, Warner and Settersten 2016, Weden et al. 2011). A methodological approach able to model complex constructs, from a latent perspective, is latent class analysis (Collins and Lanza 2010).¹ Latent class theory provides the conceptual, mathematical and statistical framework to measure categorical latent variables. Categorical latent variables can encode observed information from a set of measures, accommodating intricacy through an inductive generative process. They do so on the basis of a measurement model encompassing two or more observed categorical indicators that function to reflect an underlying grouping variable (Collins and Lanza 2010).

As distinct from the relationships that exist between a set of indicators, the substantive focus of LCA is, as such, on the within-subject patterning of assembled indicators. On this basis LCA is often referred to as a person-centred, or pattern-oriented approach. Moreover, the focus of LCA enables the identification of heterogeneity within a population; a process that facilitates the segmentation of the population into homogeneous subgroups, each having a distinctive nature that enables these segments to function as epidemiological contrasts.

Applied to the study of neighbourhoods, and within- and across neighbourhood-level typological configuration, LCA can be framed as a neighbourhood-centred approach (Warner and Settersten 2016); an approach that situates the neighbourhood unit at the forefront of both conceptual thinking and applied analyses (Warner and Settersten 2016). Within the schema of a LC-model, a neighbourhood centred approach aims to configure the clustering of neighbourhoods on the basis of a shared nature into types or classes, which are both empirically and qualitatively distinct from others in the series. Moreover, in framing neighbourhood types through a LC-approach, neighbourhoods are considered holistically, with their underlying (latent) nature reflective of the intersection of modelled features (Warner and Settersten 2016, Weden et al. 2011).

¹ In keeping with the literature, throughout this paper the phrase "latent class" is often used in a broad inclusive manner so as to succinctly encapsulate both latent class and latent transition analysis.

Measurement approaches related to and inclusive of LCA have been applied to the study of neighbourhoods (for example, McDonald et al. 2012, Palumbo et al. 2016, Wall et al. 2012). Less readily applied have been longitudinal extensions of LCA that aim to characterise, configure and model the evolution in the nature of neighbourhoods (for example, Richardson et al. 2014).² Modelling neighbourhoods over time using a discrete and categorical latent variable approach has the capacity to reveal insights able to complement existing research concerned with neighbourhood dynamics (Lekkas et al. 2017a, 2017b, Weden et al. 2011). Moreover, LCA and its longitudinal extensions, offer advantages to alternative methods that form the basis of many extant enquiries such as factor analytic methods, and cluster analysis techniques. These advantages have been outlined elsewhere (Lekkas et al. 2017a). In brief they are: insight into the manner by which a set of measures come together to reflect and characterise distinct neighbourhood archetypes while obviating the need for a series of higher order interaction terms; a measurement model, which can accommodate uncertainty and measurement error; a model-based approach to data reduction and clustering that enables the use of classical statistical criteria to assess model traits such as model fit; the probabilistic assignment of neighbourhoods – the units of analysis – to classes and dynamic profiles (latent statuses), a process that can reduce misclassification bias; and, a flexible structural model that can be applied to assess the stability or change of neighbourhood types over time, as well as mechanisms associated with the evolution of neighbourhood types, and therein neighbourhood change.

1.4 Neighbourhood-level Social Fragmentation and Social Typographies

Neighbourhoods are intrinsically about people; embodying and reflecting their social conditions of living. An integral component of these conditions are factors such as social attachment, social support, social capital, collective efficacy, social cohesion and social integration, *or* its inverse social fragmentation (Baum and Palmer 2002, Berkman et al. 2000, Ivory et al. 2012). These social factors vary by levels of magnitude in their distribution across residential spaces (Hanibuchi et al. 2012a, 2012b), and across time (Schmidt et al. 2014). They have also been tied to population-levels of well-being, quality of life, and health (Ivory et al. 2011, Jones et al. 2014, Lucumi et al. 2015). Moreover, factors such as social fragmentation are distinct from recognised neighbourhood-level traits such as material deprivation, differing: conceptually, in their measurement, in their socio-spatial patterning, and in their relationships to health and social outcomes (Congdon 2004, Fagg et al 2008, Ivory et al 2012).

² The studies by Meyer et al (2015) and Weden et al (2011) applied LCA to profile neighbourhoods at two or more time points. However, neither of these studies formally applied advanced longitudinal extensions of LCA in their analyses; rather these studies compared and contrasted cross-sectional LC-structures that were modelled across time. In comparison, Richardson et al (2014) applied the apparatus of longitudinal LCA (L-LCA) to model the neighbourhood profiles from a socioeconomic perspective across 20-years and 5-time points.

It may not always be possible to directly measure neighbourhood-levels of social fragmentation. Indirect approaches to the measurement of social fragmentation include the Congdon Index (Congdon 1996), and the New Zealand (NZ) Index of Neighbourhood Fragmentation (Ivory et al. 2012). These two indices are aggregate measures that respectively amalgamate a cogent set of census-based derived variables of population and household characteristics considered foundational to the context of social fragmentation (Ivory et al. 2012).

In formulating the NZ Index of Neighbourhood Fragmentation, Ivory et al. (2011, p1995) introduced the notion of a social topography, a concept expressed to represent the “characteristics of the group [which] act as the foundation upon which the group operates.” Applied to neighbourhood-level social fragmentation, Ivory et al. (2011, p.1995) stated that a social topography represents the set of conditions which have coalesced to establish the ground “for the level of integration and regulation within the group and therefore the level of social support available to members (probably in a recursive manner). A given social topography may be regarded as socially fragmenting if it inhibits integration and regulation in the social group.”

From a conceptual perspective, the notion of a social typography readily aligns with a neighbourhood-centred latent variable approach. To date, social fragmentation has not been explored in this manner. Moreover, no study has examined the temporal behaviour of a neighbourhood’s fragmenting conditions.

1.5 Data and Methods

The longitudinal nature of neighbourhood-level social fragmentation was modelled through the application of latent transition analysis (LTA). A general overview of the methodological approach is provided here, with the detailed methods described thereafter. In brief, the first step involved the development of a longitudinal spatio-temporal neighbourhood database. The database was formed using neighbourhood-level census data, for applicable neighbourhoods within the Adelaide metropolitan region, Australia, for the years 2001, 2006, and 2011. In the next step, at each time point, neighbourhood-level social fragmentation was measured as a latent categorical variable. This enabled the grouping of neighbourhoods with a similar nature of social fragmentation at the beginning of the study period, as well as over time. The model of change was discrete (categorical), with a discrete model of change making no assumptions as to the form of change, for example linear or quadratic (Collins and Lanza 2010); in this manner neighbourhoods may be characterised to change along a complex array of paths.

1.5.1 Study area

The study setting was the metropolitan region of Adelaide, the capital city of South Australia. Adelaide is the fifth largest city, by population, in Australia with an increase in population from some 1-million in 2001 to approximately 1.2-million in 2011 (ABS 2011a). The metropolitan boundary, delimiting the outer spatial extent of neighbourhoods encompassed within this study, was defined by the ABS Adelaide Capital City Statistical Division (ASD, Figure 1-1; ABS 2012). The ASD boundary was considered appropriate for this analysis as the focus was on urban neighbourhoods. An alternative boundary recently designated by the ABS – the Greater Adelaide Statistical Area (GASA) – is a boundary that stretches to include small-to-large peri-urban populations and townships. However, peri-urban and rural localities were considered unsuitable for these analyses given likely differences in both the levels of social fragmentation between urban and peri-urban/rural contexts (Ivory et al. 2012), and the developmental processes associated with these respective geographies (Friel et al. 2011). In addition, substantive decreases in population density are generally evident for neighbourhoods at, and beyond the margins of the ASD.

Figure 1-1 Adelaide Statistical Division (ASD)

1.5.2 Spatial - neighbourhood - unit

In the context of this study *suburbs* were operationalised as *neighbourhoods*. Australian suburbs are designated localities officially gazetted by the Geographical Place Name authority in each of the federated states and territories. While the ABS, the Statutory Authority responsible for the conduct of the national Census, does not formally recognise the suburb within its geographic framework (formerly the Australian Standard Geographic Classification (ASGC), currently the Australian Statistical Geography Standard (ASGS)), it supports a spatial approximation to these gazetted localities, this being the State Suburb (SSC; ABS 2016). For 2011 boundaries and census-based statistics, SSCs are produced via the aggregation of one or more Statistical Areas Level 1 (SA1s). Analogous to the notion of a census tract an SA1 is the smallest geographic area for which most 2011 Census data are released.³

A number of reasons underlined the selection of the suburb as a proxy for neighbourhoods. Suburbs are ubiquitous in the Australian context, forming an integral component of urban life (Davison 1995, Randolph and Freestone 2008). Beyond a spatial or locational geography, suburbs are also sociological and relational entities that evoke a sense of home, place and territory, socio-physical homogeneity, social connection, and access (Howard 2011, Simic 2008). In addition, in terms of population size, the suburbs of Adelaide approximate the scale of US-based census tracts (these ranging in size from 1200 to 8000 people; United States Census Bureau, 2001), and are congruent with the WHO's expression of a neighbourhood (WHO 2016). Furthermore, perceptions have softened with regards to the hierarchical divide that often positioned cities and their suburbs as distinct and divided geographies (Florida 2017).

1.5.3 Neighbourhood spatio-temporal database

To explore change over time at the suburb (neighbourhood) level, a longitudinal database was developed based on the SSC. For this purpose data were drawn from two sources. First, digital boundaries, in the form of ESRI Shapefiles for the 2001, 2006 and 2011 SSCs were attained from the ABS (ABS 2001a, ABS 2006b, ABS 2011b). These geographic data were complemented with property-level data (spatial boundary files encompassing all legal land parcels (dwelling points) including their attributes) attained from the South Australian Digital Cadastral Database (DCDB), for years 2001, and 2006; a database provided by the Land Services Group, South Australian Government, Department of Planning, Transport and Infrastructure.

On account of factors that include population change, the boundaries for many small-area census geographies are readily morphed between successive censuses. This impedes and challenges spatio-temporal research. To enable spatio-temporal exploration a consistent geographical approach was adopted through the harmonisation of boundaries (Norman et al. 2003). The harmonisation of geography was

³ In 2011 the ABS changed its system of census geographies on the basis of a change to its underlying spatial framework. This change saw the SA1 spatial unit replace the previous smallest area division; the census district (CD). However, in broad terms the CD and SA1 represent comparable spatial scales, each accommodating on average between 200-300 households. For 2001 and 2006 ABS generated SSC were based on the aggregation of CDs.

achieved by fixing the 2001 and 2006 SSC boundaries to the 2011 SSC boundaries. The longitudinal database then functioned to bridge data from these prior census periods to the 2011 SSC boundaries. In its approach, this method was analogous to the process used to create the US-based Longitudinal Tract Database (LTDB; Logan et al. 2014). Moreover, the advantage of adopting a forwards-oriented approach was that it captured and accounted for the manifestation of new neighbourhoods and the morphing of others.

Following on from geographic harmonisation, creation of the socio-spatial time series was based on the conceptual framework and methodology outlined by Simpson (2002). Within a Geographic Information System (ESRI ArcGIS 10.3), the first step in the process, was to overlay the 2011 SSC boundary file respectively onto the 2001 and 2006 SSC boundary files thereby linking every SSC base unit for given census years to at least one 2011 SSC (allowing for the intersection of 2001 or 2006 SSCs with multiple 2011 SSCs).

Given overlap in boundaries across years the conversion of data from one time period to another required approximation. Approximation was necessary for this task as no direct population-based measures existed to apportion data from the source to target geography on the basis of the extent of the overlap intersection (Norman 2004). For apportionment, the spatial distribution of residential dwellings was considered a suitable proxy indicator and weighting criterion (Simpson 2002). Ancillary data, attained from the DCDB, was therefore drawn into the GIS, with the spatial distribution of dwellings used to construct weights for SSC-level 'population' characteristics from prior census years within the 2011 SSC boundaries. When a SSC was completely contained within the boundaries of a single 2011 SSC a weighting of 1.0 (or 100%) was assigned to it. When a 2001 or 2006 SSC was located in more than one 2011 SSC, the ratio of its dwellings residing within each of the 2011 SSC intersected fragments formed a weight whose value was greater than 0 but less than 1, with weights for records of the same source SSC unit summing to 1. The culmination of this process was the generation of a 2001-to-2006 SSC to 2011 SSC geography conversion table (Simpson 2002) with derived SSC-level indicators estimated by using equation-1:

$$\hat{y}_t = \sum_s \frac{dwelling_{s|t}}{dwelling_s} y_s \quad [1]$$

where $\hat{y}_t$ is the estimated indicator value for an SSC, y_s is the indicator value for the source SSC, $dwelling_s$ is the number of dwellings within the source SSC, and $dwelling_{s|t}$ is the number of dwellings in the zone of intersection between the source SSC and the corresponding SSC.

While interpolation based on a combination of area and population weighting is the current standard in the field (Logan et al 2014), interpolation based on dwellings afforded more realistic weights than those assigned from simple areal interpolation as the latter assumes a uniform distribution of the data to be weighted (Goodchild and Lam 1980, Simpson 2002); an assumption that was unlikely for the set of

population attributes of interest. To the extent that dwellings are highly correlated with and therefore representative of underlying population attributes, this procedure also afforded advantages compared with weighting SSC fragments by their area alone (Gregory and Ell 2005, Simpson 2002).

On the basis that uncertainty is introduced in the conversion of data between geographic units of differing nature, data outputs for target SSC units are estimates (Simpson 2002). To ascertain the extent of estimation involved in data conversion two additional statistics were generated: the degree of hierarchy and the degree of fit (Simpson 2002). The degree of hierarchy (equation-2) reflects the percentage of source units ('s', here SSCs) with weight 'w' equal to one (i.e. 2001 or 2006 SSCs units wholly contained within a given target unit ('t', here 2011 SSCs)). Given its nature as an estimator of perfect correspondence, the degree of hierarchy "can be quite low even for geographies that are approximately equal" (Simpson 2002, p.77). Conversely, the degree of fit (equation-3) sums the maximum weight for each source unit expressed as a percentage for all source units. In this manner "the degree of fit shows more precisely the proportion of data that is not subject to estimation" and can make evident the scale of correspondence (Simpson 2002, p.77).

$$\text{Degree of hierarchy} = 100 \times \frac{\sum_{s,t}(w_{st} = 1)}{\sum_{s,t}(1)} \quad [2]$$

$$\text{Degree of fit} = 100 \times \frac{\sum_s(\max. w_{st})}{\sum_s(1)} \quad [3]$$

1.5.4 Time

A 10-year time-span was considered suitable for the study of change in urban contexts at the small area-level on the basis that it represents a sufficient interval for evolution to unfold and become evident. For example, in a US-based study across 5-years, where census tracts functioned as proxies for neighbourhoods, Mair et al. (2015) observed marked declines in neighbourhood-levels of social stress and violence, as well as increased levels of social cohesion and safety. Wight et al. (2013) analogously reported on neighbourhood-level change through the purview of US-based census tracts, documenting substantive changes in the proportion of unemployed persons, this time across a 10-year period. Ten years was also the time period used to study the spatial distribution and temporal trends in social fragmentation at a small area geographical level for the whole of England and its associated administrative regions (Grigoroglou et al. 2019). While locally, for the years from 2001-to-2006, Kupke et al. (2011) noted one in five of Adelaide's suburbs had experienced at least a 50-percent increase in the level of medium density development. And, Coffee et al. (2016) chronicled noticeable 10-year patterns of change in population density across the geographic extent of metropolitan Adelaide.

Multivariate analyses have also highlighted the propensity of neighbourhoods to exhibit multidimensional change over time periods in the range of 10-years. For example, Delmelle and Thill (2014) noted a moderate level of temporal instability in the classification of neighbourhoods assigned to three of five quality of life classes across ten-years.⁴ While, Weden et al. (2011) classified a quarter of census-tract defined neighbourhoods into a different archetype at a second time point measured ten years after the first.⁵ In addition, many neighbourhood-level intervention programs often set themselves short-to-medium term horizons to effect observable change. For example: ‘The New Deal for Communities’ programme, a United Kingdom based initiative, aimed over 10-years to positively transform deprived neighbourhoods (Lawless et al. 2010). Analogously, the US-based Building Healthy Communities initiative, is a 10-year, \$1-billion plan that aims to transform the health supporting nature of neighbourhoods (BHC Connect 2017). And, a series of area-based initiatives in Andalusia, Spain, implemented over a time frame spanning from 7-to-15 years, were designed to induce neighbourhood change for the purpose of affecting social inequities (Moya et al. 2017). While, White et al. (2017), through a natural experiment, observed impacts on mental health 7-years after a Welsh situated “Communities First” area-level regeneration program.

Although neighbourhoods *can* and *do* evolve in the short-to-medium term, neighbourhoods are often characterised by trait-based stability over such intervals. For example, in their examination of the socioeconomic status of Chicago’s neighbourhoods, across a period of twenty years, Sampson and Morenoff (2006, p.176) observed “decidedly more variation *between* neighbourhoods than there [was] change *within* neighbourhoods over time.” This durability of character was also evident across the neighbourhoods of Los Angeles, as reported in a more recent analysis that spanned an analogous period of time (Sampson et al. 2017). Moreover, for Chicago, the durable socioeconomic nature of its neighbourhoods was found to be structured and reinforced by the initial set of neighbourhood-level socioeconomic conditions (Sampson and Morenoff 2006); though such path dependence was not as evident for the neighbourhoods of Los Angeles (Sampson et al. 2017). Durability again was the predominant characteristic that arose in the ten-year analysis of the spatial patterning of social fragmentation in England as reported by Grigoroglou et al. (2019).

From an analytic perspective, invariant or rarely changing characteristics, such as a (near) stable long-run neighbourhood trait, may be considered of peripheral interest, or as a statistical nuisance (Bell and Jones 2015, Gunasekara et al. 2014). This despite the capacity for (near) unchanging, time-invariant variables and processes to effect time-varying variables and processes (Bell and Jones 2015); a facet exemplified by the

⁴ In the study by Delmelle and Thill (2014) neighbourhoods were defined on the basis of Neighbourhood Statistical Areas. Moreover, neighbourhoods were assigned to a quality of life class based on a composite index composed of 17-indicators.

⁵ In the study by Weden et al (2011) neighbourhoods were measured using a set of some 32-indicators capturing features of the: built environment, migration and commuting behaviours, socioeconomic composition, and demographic and household composition.

correspondence between stable political democracies and their state of peace across time (Bell and Jones 2015). And, it equally may be that for a neighbourhood-level characteristics such as social fragmentation, that stability, for example, at a low level, is requisite for the staging of social integration and regulation.

1.5.5 Measures

Data for the measurement of neighbourhood-level social fragmentation were sourced from the national Census of Population and Housing, Australian Bureau of Statistics (ABS), for years 2001, 2006 and 2011 (ABS 2001b, 2006a, 2011c). The five-yearly national census provides rich information on the characteristics of the Australian people and their dwellings, with data provided at a number of geographic levels.

For given census years, the ABS provides two population counts for census data; these respectively are based on the enumerated and usual place of residence. The enumerated count encompasses the number of people usually resident or transient (for example visitors, night-shift workers) in a particular area on the given Census night. In contrast, the usual residence count reflects the characteristics of residents irrespective of where those residents were on the Census night. For this study, with few exceptions, the indicators used to reflect neighbourhood-level social fragmentation were derived from the usual place of residence count.

1.5.5.1 Social fragmentation

Latent class theory suggests there are underlying, or hidden, subgroups in a population. This heterogeneity cannot be directly observed but must be inferred from a set of categorical items. In this study the measurement model for neighbourhood-level social fragmentation was based on a set of nine census-derived indicators, namely: home ownership, residential stability, single-person households, married persons, non-family households, children, migrants, non-English language speakers, and long-term residents. These nine indicators of area-level social fragmentation stem from the work of Ivory et al. (2011, 2012). The indicators capture three dimensions which inter-relate, and are theorised to “‘fragment’ neighbourhood collective social functioning” (Ivory et al. 2012, p.975). The three dimensions represented by Ivory et al.’s (2012) nine measures are: norms and values, social resources and attachment. Moreover, these dimensions represent: the ‘how’ (limited means of communicating norms and values across neighbourhoods); the ‘who’ (neighbourhood social resources); and the ‘why’ (levels of attachment to people and place) of fragmenting social (neighbourhood) conditions (Ivory et al. 2012).

Within the context of this study the nine indicators were operationalised as follows: home ownership - *place of usual residence owned outright, with a mortgage or on a rent to buy scheme*; residential stability - *persons who had been resident at their usual address as per one year ago*; single-person households - *a one person residence*; married persons - *couples in a registered marriage or in an identified de facto relationship inclusive of lesbian, gay, bisexual, transgender, queer, and intersex couples*; non-family households - *non-related persons resident in a ‘group’ household or share house (i.e. flatmates)*; children - *persons aged 14-*

years and younger; recent immigrants - persons born overseas who had arrived in Australia less than five years ago; non-English language speaking proficiency - residents who do not speak English well or at all; and, long-term residents⁶ - persons who had been resident at their usual address as per five years ago.

Each of the nine census indicators was initially calculated as a proportion based on the underlying population in each neighbourhood unit at each of the time points, with 2001 and 2006 indicators subsequently weighted to enable spatio-temporal concordance as previously discussed (section 5.5.3). Thereafter, for LC-analyses all proportions were dichotomised via median splits that were time-varying in nature. Dichotomisation was undertaken for two reasons: concerns related to the over-extraction of latent classes when latent classes are enumerated with their indicator variables in continuous form (Bray B. per comms. 2015);⁷ and, the distributional properties of a number of the indicators, specifically excessive kurtosis (see table-1.2).

While mitigating some issues, there was however cognisance that dichotomisation could introduce others; namely information loss through a reduction in heterogeneity (MacCallum et al. 2002), and therein the identification of incomplete subgroups. In addition, there was awareness of dichotomisation potentially impacting the meaningfulness of identified classes through its effects on within-class homogeneity and across-class heterogeneity where values at the extremes of their associated indicator distributions were treated equally with those just above or below their respective median splits. Recent simulation analyses, albeit involving latent classes reflective of count-based indicators, however suggests that such concerns may be unfounded. Specifically, in circumstances where the number of estimated classes was greater than the number of indicators in the measurement model, median split dichotomisation did not affect the quality of mixture recovery, or bias parameter estimates (Macia and Wickham 2018).

From an applied perspective consideration has previously been afforded to the impact of dichotomisation in LC-models. Notably Weden et al. (2011, 2010) examined the latent structure of a sample of US-based census tracts across two time points, 10-years apart, with latent classes identified using some 32 measures which were primarily dichotomised using median splits. Of relevance here, Weden et al. (2011, 2010) conducted a sensitivity analysis comparing modelled heterogeneity against a scenario where the indicators were applied in their original continuous-variable form. On the basis of their analysis the latent class

⁶ In the NZ Social Fragmentation index “long-term residents” were operationalised on the basis of continuous residence over a period of 15-years. The ABS however only provides data pertaining to residential mobility for 1- and 5-years.

⁷ Bray (2015, per comms.): “Regarding issues of model identification, selection, and interpretation there are a number of idiosyncratic issues that arise when using Latent Profile Analysis (LPA), mostly due to the nature of mixtures of normal distributions being unbounded. LPAs can be difficult to identify (in terms of the ML-solution), particularly when item variances are not restricted to be equal across classes, which is an assumption that is sometimes not palatable; they often have poorer model identification than LCAs with the same number of indicators. Models can also often be difficult to select because the AIC, BIC and other fit criteria often continue to prefer larger and larger models; they often have poorer entropy than LCAs with the same number of indicators. In terms of interpretation, the concept of homogeneity is not the same as in LCA because the item means are bounded by the scales used (not by 0 and 1), and the concept of separation can be difficult to assess if item variances are relatively large.”

solution using median splits was reported as robust to the alternate ‘continuous indicator’ specification (Weden et al. 2010). Moreover, the authors reported that the “distribution of census tracts across groups was more skewed with the cluster analysis” using continuous indicators as compared with the dichotomised approach (Weden et al. 2010, p.44).

1.5.6 Statistical analyses

As expressed, discrete change in neighbourhood types over time was modelled using LTA. A general description is first provided of LTA followed by the process of model identification. A representative path diagram for the LTA is depicted in Figure 1-2. Further methods-oriented details are provided in Appendix-A, these encompassing: the mathematics of LTA, the estimation process involved with LTA, and the assignment of the units of analysis to latent classes. For completeness, Appendix-A also presents information on issues not applicable to analyses presented in this paper, namely processes related to missing data considerations in the schema of mixture models, and approaches to covariate analysis (1-step compared with 3-step approaches, inclusive of bias-adjustment).

Latent transition analysis models change as transitions in latent states across time; in this study, latent neighbourhood-levels of social fragmentation across three time points. Change is represented by a transition probability matrix. Each element of the matrix identifies the probability of transitioning to a latent state at time $t+1$ given membership at time t . Change is autoregressive and represented by discrete movement - state changes - conventionally modelled across consecutive time points through an autoregressive-1 (AR1) specification, though it is possible to model more complex scenarios, for example and AR2 model where $t3$ is regressed on $t1$ (Collins and Lanza 2010).⁸

Figure 1-2 LTA path diagram (unconditional AR1 model with 3-time points and 9-indicators)

⁸ For pragmatic purposes all analyses were restricted to AR1 models.

There are systematic, stepwise procedures to guide the conduct of a LTA. Procedurally, this study followed the recommendations articulated by Collins and Lanza (2010), Masyn (2013), and Nylund (2007). Furthermore, in the context of exploratory model building, these step-wise processes function to enhance transparency of model-selection. While step-wise, it is important to note that the identification of latent class models is also an iterative process, with consideration given as much to the meaningfulness of enumerated classes as it is to statistical estimates. The following sequential steps were followed: estimate and examine a k -series of cross-sectional LCA measurement models independently at each of the three time points so as to identify an optimal (well-fitted) k -class model; using the best-fitting k -class LCA model, explore estimation of an AR1-LTA model, first imposing measurement invariance across time, and then freely estimating the LTA model comparing model fit across these two variants;⁹ finally, explore stationarity of the propensity for neighbourhoods to transition from one class to another across the two respective intervals i.e. probability of transitioning between 2001-to-2006 as compared with 2006-to-2011.

1.5.6.1 Model parameters

Latent transition analysis involved the estimation of three sets of parameters. These parameters were: a vector of dynamic class membership probabilities at Time 1 (δ) that specified the time-specific prevalence of each class in the population; matrices of transition probabilities (τ), these reflecting the incidence of neighbourhood-level transitions in social fragmentation from Time 1 to Time 2, and Time 2 to Time 3; and, a matrix of conditional (class-specific) item-response probabilities for each of the indicators in the measurement model at each point in time (ρ).¹⁰

1.5.6.2 Analytic process: model assessment and selection

Following measurement specification, the process of enumerating a latent class model proceeded by running a $k=1$ class solution, then specifying an increasing series of $k+1$ unconditional models (Collins and Lanza 2010), with an upper bound of $k=9$. To enable model assessment, *model fit* was monitored (Collins and Lanza 2010). Absolute, overall goodness-of-fit measures, for example the likelihood ratio chi-squared statistic (G^2 ; G-squared), can be problematic with latent class models, in particular longitudinal applications, where the contingency table may be demarcated with many sparse cells.¹¹ This sparsity sees the G^2 statistic no longer possessive of a theoretical chi-squared distribution (Masyn 2013). **Relative fit measures** can be

⁹ Imposing measurement invariance refers to the process of fixing or constraining the class-specific item-response probabilities to be equal across time. This serves two purposes. First, the qualitative interpretation, or meaning of the latent statuses is maintained across time which aids interpretation. Second, estimation is facilitated on the basis of holding the set of item-response probabilities constant across time. The inverse of imposing measurement invariance is to 'freely estimate' the latent statuses across time by no restricting the estimation of these model parameters (Collins and Lanza 2010).

¹⁰ With regards to LTA, within the literature, classes that are identified are often referred to as 'latent statuses', or 'latent profiles' so as to distinguish them from the latent classes generated in cross-sectional LC-models.

¹¹ For an LTA model with 9-dichotomous indicators measured at 3-time points there are $(2^9)^3$ or $2^9 \times 3 = 134,217,728$ possible response patterns (W).

helpful alternatives, noting though that “even if one model is a far better fit to the data than another, *both* could be poor in overall goodness of fit” (Masyn 2013, p.567). But which measures of fit?

Within the literature there is no universal agreement on the set of tests or criteria to apply in order to guide the identification of a good-fitting latent class measurement model (Nylund et al. 2007, Nylund-Gibson and Masyn 2016). Best practice recommendations are to form judgments derived from multiple vantages, including: the application of the **Lo-Mendell-Rubin likelihood ratio test** (LMR-LRT), the **Bootstrap Likelihood Ratio Test** (BLRT), several **information criteria** (Bayesian and sample-sized adjusted Bayesian Information Criteria (BIC and α -BIC), the **interpretability** and **meaningfulness** of competing solutions, **class proportions**, and **entropy** (Collins and Lanza 2010, Masyn 2013).

The **LMR-LRT** and **BLRT** compare successive latent class models to one another (for example k vs. $k+1$ class model). Moreover, both the LMR-LRT and BLRT “provide a p -value which is used to indicate whether a k -class model fits the data statistically significantly better than the $k-1$ class model” (Nylund-Gibson and Masyn 2016, p.789).¹² Of note, contrary to simulation analyses (Nylund et al. et al 2007), substantive disparities have been reported in real-world analyses in terms of size of the p -values respectively generated from the LMR and BLRT (Muthén 2009). In such circumstances advice is to defer to examining the patterning of the information criteria (Muthén 2009). The **BIC** and the sample sized adjusted **α -BIC** describe parsimony of the model, the smaller the coefficient, the better fitting and more parsimonious the model. These information criteria were selected based on their performance in simulated mixture model analyses (Nylund-Gibson and Masyn 2016).

Interpretable latent class typologies are denoted by a classificatory structure that captures the multidimensional patterning of the construct of interest, as well as differentiating between different types of members. With regards to the interpretability of competing model solutions, important considerations are the magnitude of **latent class separation and homogeneity**; with high levels of both desired (Lubke and Neale 2006). Latent class separation relates to the manner in which each level of the latent variable characterises for example, a distinct mixture of social fragmentation. While latent class homogeneity relates to the unique correspondence between each latent class and the observed indicator items (Collins and Lanza 2010).

¹² The LMR-LRT uses an analytic approximation for the sampling distribution of the log-likelihood ratio test statistic (Lo Y, Mendell NR, and Rubin DB (2001). Testing the number of components in a normal mixture. *Biometrika* 88(3): 767-778.; The BLRT generates an empirical approximation for the sampling distribution of the log-likelihood ratio test statistic using a parametric bootstrapping method (McLachlan (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 36(3): 318-324.)

Homogeneity and separation were examined empirically by attending to the values and patterning of the item-response probabilities, within- and across classes (the ρ -parameters, which within a given class should exhibit probabilities close to 0 or 1 (Wurpts and Geiser 2014)). Homogeneity and separation were also considered qualitatively, by asking reflective questions of the nature: *Collectively, are the within-class patterns of item-response probabilities measuring something distinct? And, are the within-class patterns of the item-response probabilities able to explain distinctions across projected levels of the latent variable?*

A further statistic considered to adjudicate a model's meaningfulness was the estimated time-specific **proportions of each class** in the population (the δ -parameters). Candidate models with classes displaying a relatively low prevalence may be indicative of less than meaningful models, this despite empirical evidence suggesting they are the best fit to the data. There is no minimum threshold relating to class prevalence however, models with classes comprising less than 1-5% of the total sample warrant greater scrutiny (Berlin et al 2014).¹³

Finally, **entropy** was also an important concern. In the context of latent class models, entropy relates to the precision of classification, or the uncertainty of assigning the units of analysis (here, neighbourhoods) to classes. Assignment was based on the set of subject- or unit-specific posterior probabilities. Posterior probabilities reflect each subject's or units' observed pattern of responses, conditioned by class, and calculated using an application of Bayes theorem. As true class membership is unknown, there is classification error (Bolck et al 2004). Uncertainty of classification is high when the set of posterior probabilities for given units are similar across classes. Classification uncertainty can be measured in a number of ways. A normalised, scaled version of entropy at the global level (across all classes) was derived using equations-7 and 8, where $\hat{\alpha}_{ic}$ is the estimated probability that observation i is a member of class c :

$$EN(\alpha) = - \sum_{i=1}^N \sum_{j=1}^J a_{ij} \log \alpha_{ij} \quad [7]$$

and,

$$E = 1 - \frac{EN(\alpha)}{N \log J} \quad [8]$$

¹³ In relation to a minimum class prevalence, Frankfurt et al (2016, p.634) cite the guidance provided by Jung & Wickrama (2008) who "suggested that at least 1% of the total sample should comprise each class; however, this rule depends on the sample size. For example, in a large sample, 1% may describe hundreds of [neighbourhoods] and thus be a meaningful class. In a smaller sample, 1% of the sample may be too small to define a class."

, with the values for entropy ranging from 0-to-1. As a general guide, entropy values nearer one indicate improved enumeration accuracy and classification certainty (Celeux and Soromenho 1996). A 'local' or class-specific measure of entropy was also considered by calculating the set of average within-class posterior probabilities after units (subjects) have been assigned to their most likely class. A local measure of entropy functions to complement the global entropy metric, and may highlight classes with high levels of classification uncertainty that are otherwise masked by the global metric. Again, values near one are indicative of higher classification certainty, with a rule-of-thumb expectation for meaningful models being they exhibit averages that exceed 0.70 (Nagin 2005).

After initial latent class enumeration, the process shifted to a consideration of the key model assumptions and restrictions (Nylund-Gibson and Masyn 2016). An integral assumption of latent class models is *local independence*. Local independence denotes that within a given class, conditional on the level of the latent variable, the observed indicators are independent. As multiple indicators ($n=9$) were applied in this study to measure social fragmentation, the assumption of local independence may be considered unlikely, particularly given known processes associated with neighbourhood selection that might have seen people and households attracted by, and or self-selecting into neighbourhoods (Bruch and Mare 2006, Sampson and Sharkey 2008).

Local independence was evaluated by inspecting the bivariate residuals between pairs of indicators (Reboussin et al. 2008). The bivariate residuals are standardized Pearson residuals. In general, the applied rule of thumb is that any residual larger than 1.96, in absolute terms, is significant where $\alpha = 0.05$ (Nylund 2007). And, better fitting models are those with few large standardized residuals, and a relative low overall percentage of significant residuals (not in excess of 5%; Masyn 2013, Nylund 2007).

To examine the transition patterns, the conditional item-response probabilities (matrix of p parameters) were constrained as invariant across all time points. Imposing *measurement invariance* fixes the nature of the classes across times, a process that facilitates the interpretation of the transition probabilities. Imposing measurement invariance also functions to reduce the number of parameters estimated. However, from a developmental perspective, the assumption of measurement invariance may be unduly restrictive where there might be reason to perceive time-varying evolution in the latent structure.

As models with different levels of measurement invariance are nested, the applicability of measurement invariance was assessed using the likelihood-ratio chi-square difference statistic - G^2 (*diff.*) (Lanza and Bray 2010). Concurrently invariance was assessed with reference to the respective set of information criteria (i.e. AIC, BIC, a-BIC), and the nature of generated latent status profile plots for interpretability and comparability. The same process of nested model comparison was undertaken to explore the applicability of *stationarity* in the transition probabilities (i.e. *transition probability invariance*) across the two time periods of change.

Graphics, or visualisation, are recommended to complement numeric output and facilitate model identification and interpretation (meaningfulness). Notably, **item response plots** were generated to display the within-class item-response probabilities (p parameters), visually summarising the pattern of response probabilities for each latent class and aided in the consideration of homogeneity. Item response plots were also considered in the context of examining measurement invariance. As the units of analysis in this study are neighbourhoods, the geo-visual capacity of ArcGIS was called on to project through **chloropleth maps** the spatio-temporal distribution of neighbourhoods modally assigned to categories of social fragmentation. With regards to model interpretation and identification the utility of this mapping process lies in its capacity to visually consider the manner in which likely latent class solutions map onto geo-spatial information characteristic of local neighbourhoods. Furthermore, as a visual language, mapping can induce relational forms of sensing and understanding (Dovey and Ristic 2017), that can complement the numeric information generated by latent variable models for the purposes of latent class enumeration.

Latent transition analysis was performed with **maximum likelihood estimation**, using an expectation-maximization (EM) algorithm, and modelled with robust standard errors (MLR). To guard against the identification local maxima, a minimum of 1000-sets of random start values was computed (Geiser 2013, Nylund-Gibson and Masyn 2016). In conjunction, 100 of the starting values that yielded the largest log likelihood values in the first step of the EM-process were specified to carry over into the second step of this algorithmic optimization (Geiser 2013). The number and proportion of this 100 set of random starting values that converged to a proper solution were recorded; with the replication of the apparent global solution across random starting values functioning as a guide to confidence in the identification of the best or 'true' ML solution (Masyn 2013). Where the best ML solution was not readily discernible from the next best log-likelihood values, competing models were rerun using seeded starting values with consistency of the latent structures in the various iterations serving to guide decision making (Hipp and Bauer 2006). In addition, the number of random starts was doubled to $n=2000$. Syntax for all procedures is provided in Appendix-B. Significance tests were based on $\alpha = 0.05$. Relevant coefficients and confidence intervals are reported for all results (for example fit indices, χ^2 -diff, p -values, ORs and 95%-CIs). There were no missing data. Models were estimated with Mplus version 8.0 (Muthén and Muthén 2017).

1.6 Results

1.6.1 Spatiotemporal database

Table 1-1 presents statistics for the degree of hierarchy and degree of fit which reflect the extent of data conversion for 2001 and 2006 SSC-level indicators to 2011-bounded SSCs. In addition, for applicable SSCs. Table 1-1 also highlights the per cent relative change in their square kilometre area across census years.

Table 1-1 Degree of hierarchy and fit

Source geography (number of units)	Target geography (number of units)	Degree of hierarchy	Degree of fit
SSC 2001 (<i>n</i> :373)	SSC 2011 (<i>n</i> :409)	55.11 %	96.97 %
SSC 2006 (<i>n</i> :399)	SSC 2011 (<i>n</i> :409)	53.79 %	96.00 %
Relative change in area for SSCs across census periods *		Median (sq.km)	Range (sq.km)
SSC 2001 ... 2006		0.01 %	- 98.43 ... 249.05 %
SSC 2001 ... 2011		-0.05 %	- 90.70 ... 413.89 %
SSC 2006 ... 2011		-0.07 %	- 85.91 ... 5026.76 %
* Relative changes in area across census periods are based on the following sample sizes denominators: 2001-to-2006 = 366; 2001-to-2011 = 368 and 2006-to-2011 = 392, where denominators were based on SSCs being comparable by name across time (i.e. excludes new SSCs from the later time points in head-to-head comparisons)			

1.6.2 Descriptive information

Table 1-2 presents descriptive information on the social fragmentation indicators at each of the census dates. Of note, the presented summary statistics (and visualisations; Appendix C) were derived using the indicators in their continuous form. None of the indicators exhibited substantive skew, though kurtosis (exceeding an absolute value of 10; Kline 2011) was a feature of a number of the distributions. Across time, most of the indicators exhibited fairly modest upward or downward change. Though more substantive change was evident for 'Recent Immigrants'; this increasing non-linearly.

Table 1-2 Descriptive statistics for the 9-indicators of social fragmentation

2001	% Child. & Youth (< 15-yrs)	% Recent Immigrants	% Non-Eng. Speakers (Proficiency)	% Married Persons	% Home Owners	% Same Usual Residence: 1yr	% Single Person Households	% Non -Family Households	% Same Usual Residence: 5yrs
Mean	17.63	7.72	8.84	55.67	67.73	77.57	28.12	3.46	55.65
Median	17.18	6.99	6.22	56.34	69.99	79.82	29.46	2.82	57.28
SD	5.27	4.59	8.33	13.21	18.80	15.680	11.98	2.67	13.65
Min	0.53	0.00	0.00	1.84	2.34	2.29	0.51	0.00	1.71
Max	43.16	22.88	39.75	131.44	151.28	165.93	106.32	18.32	125.83
Skew	0.36	0.92	1.36	0.09	-0.16	-0.86	0.63	1.96	-0.55
Kurtosis	2.36	0.94	1.55	6.85	2.49	12.04	4.02	5.27	6.58
2006									
Mean	17.21	11.41	9.23	56.49	69.47	81.25	28.84	3.42	58.53
Median	16.84	10.63	6.47	56.46	69.35	81.64	29.15	2.58	59.74
SD	4.72	7.04	9.39	12.09	18.16	13.72	10.83	3.05	12.42
Min	3.35	0.00	0.00	9.52	12.12	15.05	3.24	0.00	10.95
Max	44.32	46.49	58.97	123.11	173.00	169.73	83.02	23.09	131.36
Skew	0.85	1.21	1.84	1.26	1.26	1.42	0.31	2.70	0.56
Kurtosis	3.99	2.92	4.56	8.03	6.63	16.22	1.26	11.31	7.94
2011									
Mean	16.95	19.31	8.94	56.47	68.50	82.58	28.11	3.96	59.51
Median	16.98	18.43	7.93	56.81	68.20	83.29	28.81	3.10	60.35
SD	3.48	10.11	7.371	8.06	14.11	5.47	8.87	3.15	9.48
Min	4.49	0.00	0.00	33.91	27.97	54.98	6.15	0.00	19.65
Max	30.41	52.95	46.94	75.51	100.00	93.84	51.33	27.14	80.08
Skew	0.14	0.52	1.24	-0.11	-0.18	-1.17	-0.12	2.70	-0.92
Kurtosis	1.33	0.23	2.16	-0.412	-0.52	3.13	-0.45	13.50	1.71
Notes: For 2001 and 2006 indicators, percentages may exceed 100 as a consequence of weighting the data for spatio-temporal concordance/harmonisation. Abbreviations: Child. = Children; Eng. = English; Prof. = Proficiency; Reside.= Residence									

1.6.3 LCA model estimates and fit statistics and plots

Nine LCA models were fitted at each of the three measurement occasions ($k=1$ through to $k=9$). These models were only estimated using information from neighbourhoods which had a minimum dwelling count of $n=35$ at each of the three time points ($n=371$). Table 1-3 presents the estimated log-likelihood values, the frequency with which models replicated the lowest log-likelihood, and estimates for the model fit indices across the time varying k -class solutions. Table 1-3 also presents the estimated class proportions, global and local entropy, frequency of imposed 'boundary estimates',¹⁴ and programmatic error messages.

¹⁴ A boundary estimate arises where the conditional item-response probability approximates the boundary of the probability distribution. Here the corresponding logit value is positively or negatively large which in the context of ML-estimation may

As ' k ' (the number of classes estimated) was incrementally increased, estimated models took on a more 'complex structure'. That is classes were generated that had an unequal prevalence. Moreover, as $k \geq 6$ the additional classes that were estimated displayed relatively small proportions ($\leq 5\%$). These smaller classes were also characterised by conditional item-response probabilities that were less likely to be near the boundaries of the probability space for all but a few of the nine measurement items; a feature suggestive of a progressively weaker conditional relationship between these indicators and the modelled latent class. While, across model iterations, entropy - global and local – was consistently high, generally exceeding 0.90. However, the frequency with which the lowest log-likelihood value was replicated diminished rapidly as the number of classes estimated exceeded six. Furthermore, the frequency with which the modelling program 'stepped in' to fix boundary estimates to enable model identification also steadily increased where $k > 5$.

In terms of the information criteria, across time the BIC consistently pointed towards a four class solution. While, the AIC and aBIC were variously suggestive of a seven and an eight class model. A series of scree plots examining the respective profiles of the AIC, BIC and aBIC estimates generated at each time point for $k=1$ through to $k=9$ models is presented in Figure 1-3. As is evident, as the number of estimated classes was increased beyond five, declines in the AIC and aBIC became relatively marginal. LMR and BLRT comparative tests between incrementally increasing k and $k+1$ solutions also presented discordant though somewhat unstable perspectives (Table 1-4). Notably, the BLRT tests failed to rule out larger models until $k=8$, whereas the LMR tests suggestively pointed towards both smaller, and larger models; these varying depending on the time period under question.

Collectively, a picture emerged suggesting an appropriate model solution was in the range from $k=4$ to $k=6$. To further study these prospective solutions, conditional item-response probability plots were generated for $k=4, 5$, and 6 solutions that were freely estimated within and across time (Appendix-D). A series of corresponding choropleth maps were also produced by 'hard classifying'¹⁵ neighbourhoods at each time point through modal assignment into social fragmentation classes for $k=4, 5$ and 6 solutions (Appendix-E).

Inspection of the conditional item-response plots from the competing $k=4, 5$ and 6 solutions afforded a number of insights, namely: the four classes identified in the $k=4$ solutions were relatively stable; both as k -increased, and across time; the additional classes which were modelled i.e. the fifth and sixth classes, appeared to be adequately differentiated from each other, as well as from the other four classes; *and*, they also featured potentially unique differences in the nature of their within class item-response patterns (i.e.

cause singularity of the information matrix. To avoid this scenario, modelling programs such as Mplus fix such estimates. A high frequency of imposed 'boundary estimates' may be a sign of a local maximum and or of model overfit (Geiser 2013).

¹⁵ In general, for each unit of analysis in an LCA and or LTA model, there is a non-zero posterior probability of membership (assignment) into each of the estimated k -classes (or, latent statuses). Moreover, these posterior probabilities sum to one across the k -classes (or, latent statuses). Units of analysis can be assigned to classes based on the highest, or modal value, of their respective estimated posterior probabilities; a process referred to as 'hard classification or partitioning'. Alternatively, a form of soft partitioning might be applied based, for example, on proportional assignment into each of the k -class where the set of posterior probabilities are applied as weights in prospective classify-analyse analyses (Bakk et al. 2014).

they displayed moderate-to-high class homogeneity). Spatially, the assessment of whether an additional fifth, or fifth and sixth class, augmented any final solution was hampered by the relatively low prevalence of these two classes. Visually, though it appeared that neighbourhoods assigned to these two classes occupied territory at the margins of the space populated by the four classes modelled in the $k=4$ solution. While this observation potentially may have been indicative of meaningful 'boundary' or transitory classes; it might also have been an artefact of the low prevalence of these classes.

Considered overall, and ceding that the BIC often tends to underfit models (Dziak et al. 2019), a four-class LCA model was selected to carry forward into the LTA. With regards to the modelled four classes of social fragmentation, these were labelled as: '**low**' (class-A; reference class in analyses), '**mixed-level inner urban**' (class-B), '**mixed-level peri-urban**' (class-C), and '**high** social fragmentation' (class-D). Coarse-grained labels were applied to avoid the reification of categorical labels, as well as to avoid mischaracterising the classes as immutable. Labelling was however based on the patterning of the conditional item-response probabilities (Figure 1-4), and the spatial patterning of neighbourhoods to assigned classes (Figure 1-5).

From a measurement perspective classes-A and -D were orthogonal to each other, being respectively characterised by the lowest and highest levels of social fragmentation across all three dimensions. Neighbourhoods modally assigned to classes-A and -D also differed in their spatial distribution; the former being prominent among the middle-to-outer ring of metropolitan suburbs, while the latter were more likely to be represented within the inner ring of suburbs. Quantitatively, classes-B and -C were also relatively orthogonal to one another, but they differed to classes-A and -D through their propensity to *not* reflect all nine measurement items at the polarities of the modelled probability distributions of these indicators. Moreover, classes-B and -C were variegated in the manner in which they respectively reflected the dimension of 'attachment'. Namely, class-B suburbs exhibited a greater propensity to harbour longer term residents *but* also unconventional and single-person household types. The opposite was the case for class-C suburbs. And, spatially, while some neighbourhoods assigned to class-C were represented within the inner and middle ring of suburbs (not dissimilar to the spatial distribution of class-D suburbs), class-C suburbs were more likely to be located at the peri-urban fringe. In terms of their prevalence, neighbourhoods assigned to classes-A and -D were the most prevalent at each measurement occasion (estimated time-varying prevalence estimates respectively ranged from 44% (2001) to 35% (2011), and 31% (2001) to 35% (2011). While, neighbourhoods assigned to classes-B and -C were the least represented; their prevalence estimates respectively ranged from some 14% (2001) to 12% in 2011, and 12% (2001) to 17% (2011).

Table 1-3 Fit statistics derived from LC-models for $k=1$ to $k=9$ solutions; 2001-2006-2011

Model statistics	1-class	2-class	3-class	4-class	5-class	6-class	7-class	8-class	9-class
LCA 2001									
Log-likelihood	-2314.406	-1937.323	-1853.080	-1797.935	-1768.710	-1748.179	-1732.011	-1716.515	-1707.563
n LL replications (max 50)	100	100	100	100	63	15	7	8	2
AIC	4646.813	3912.645	3764.160	3673.386	3635.419	3614.358	3602.023	3591.031	3593.127
BIC	4682.058	3987.053	3877.729	3826.117	3827.313	3845.414	3872.240	3900.411	3941.669
SSA-BIC	4653.504	3926.772	3785.722	3702.383	3671.852	3658.226	3653.326	3649.769	3659.300
Entropy – Global	-	0.933	0.932	0.906	0.897	0.896	0.906	0.927	0.926
Entropy – Local	-	0.983	0.969	0.943	0.947	0.922	0.928	0.932	0.945
Class size proportions	1.00	0.528	0.461	0.437	0.348	0.297	0.305	0.313	0.313
		0.472	0.412	0.288	0.286	0.288	0.253	0.251	0.237
			0.127	0.148	0.148	0.132	0.121	0.121	0.086
				0.127	0.113	0.108	0.119	0.086	0.075
					0.105	0.105	0.089	0.067	0.067
						0.070	0.070	0.067	0.067
							0.043	0.051	0.057
								0.043	0.051
									0.046
n Fixed-boundary estimates	-	-	1	1	2	6	13	18	25
Mplus error messages	-	-	-	-	-	-	-	-	-
LCA 2006									
Log-likelihood	-2314.406	-1854.761	-1775.771	-1728.157	-1699.390	-1678.991	-1662.510	-1648.221	-1638.120
n LL replications (max 100)	100	100	99	100	55	13	3	4	0
AIC	4646.813	3747.523	3609.541	3534.315	3496.781	3475.983	3463.020	3454.441	3454.240
BIC	4682.058	3821.931	3723.111	3687.047	3688.675	3707.039	3733.238	3763.821	3802.782
SSA-BIC	4653.504	3761.650	3631.103	3563.312	3533.214	3519.851	3514.323	3513.180	3520.414
Entropy – Global	-	0.923	0.887	0.882	0.903	0.936	0.933	0.933	0.909
Entropy – Local	-	0.978	0.952	0.933	0.932	0.930	0.937	0.937	0.923
Class size proportions	1.00	0.502	0.404	0.369	0.367	0.345	0.331	0.326	0.253
		0.498	0.375	0.272	0.267	0.261	0.256	0.261	0.208
			0.221	0.199	0.140	0.132	0.132	0.097	0.135
				0.159	0.129	0.094	0.094	0.081	0.094
					0.097	0.094	0.073	0.078	0.084
						0.073	0.070	0.067	0.070
							0.043	0.057	0.067
								0.032	0.056
									0.032
n Fixed-boundary estimates	-	-	-	3	4	8	11	18	22
Mplus error messages	-	-	-	-	-	-	-	-	a
LCA 2011									
Log-likelihood	-2314.406	-1728.428	-1643.576	-1598.097	-1571.999	-1554.051	-1539.669	-1527.381	-1518.53
n LL replications (max 100)	100	100	100	100	99	81	15	11	7
AIC	4646.813	3494.857	3345.152	3274.195	3241.999	3226.102	3217.338	3212.762	3214.306
BIC	4682.058	3569.265	3458.722	3426.927	3433.893	3457.158	3487.556	3522.142	3562.848
SSA-BIC	4653.504	3508.984	3366.714	3303.192	3278.431	3269.970	3268.641	3271.500	3280.480
Entropy – Global	-	0.919	0.867	0.886	0.916	0.923	0.940	0.943	0.922
Entropy – Local	-	0.978	0.939	0.934	0.939	0.932	0.942	0.943	0.945
Class size proportions	1.00	0.520	0.364	0.361	0.356	0.358	0.361	0.356	0.337
		0.480	0.280	0.350	0.345	0.332	0.334	0.337	0.329
			0.356	0.137	0.146	0.108	0.105	0.059	0.059
				0.151	0.099	0.100	0.054	0.057	0.057
					0.054	0.054	0.051	0.048	0.048
						0.048	0.048	0.048	0.048
							0.048	0.048	0.048
								0.046	0.043
									0.030
n Fixed-boundary estimates	-	-	2	1	5	8	13	15	22
Mplus error messages	-	-	-	-	-	-	-	-	-

AIC = Akaike Information Criteria; BIC = Bayesian Information Criteria; SSA-BIC = Sample Size Adjusted BIC; Local entropy = average of classification probabilities for the most likely LC-membership; Class size proportions based on estimated most likely LC-membership (listed highest to lowest by probabilities and not by subjective type (i.e. nature of classes not held constant as k increases))

a = the chi-square test cannot be computed because the frequency table for the latent class indicator model part is too large.

blue font represents a minima in the information criterion

y-axis = AIC, BIC and SSA-BIC values
x-axis = number of estimated LCA k -classes

year = 2001 = Top Left
year = 2006 = Top Right
year = 2011 = Bottom Left

Figure 1-3 Scree plots for AIC, BIC & SSA-BIC values for $k=1$ to $k=9$ LC-models (yrs: 2001-to-2011)

Table 1-4 LMR and BLRT p -values for $k \nu k+1$ LCA model fit comparisons by year

	2001		2006		2011	
	LMR	BLRT	LMR	BLRT	LMR	BLRT
1 ν 2-class	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2 ν 3-class	0.0000	0.0000	0.0016	0.0000	0.0001	0.0000
3 ν 4-class	0.0356	0.0000	0.2369	0.0000	0.0029	0.0000
4 ν 5-class	0.1276	0.0000	0.0085	0.0000	0.1398	0.0000
5 ν 6-class	0.0736	0.0000	0.1771	0.0000	0.0065	0.0000
6 ν 7-class	0.0274	0.0080	0.1668	0.0030	0.0233	0.0120
7 ν 8-class	0.0166	0.0020	0.0683	0.0090	0.2202	0.0400
8 ν 9 class	0.0206	0.2460	0.5885	0.1080	0.0191	0.1380

*Grey-shaded p -values < 0.05

Figure 1-4 Conditional item response probability plots for $k=4$ 2001-LCA solution

Spatial distribution of suburbs modally assigned (hard classified) to the x4-classes of social fragmentation at 2001, 2004 and 2001 (left-to-right) as independently estimated using LCAs at each time point. Latent classes as such were freely estimated across time however the nature of the item-response probabilities were tracked and corresponding classes across time are represented on the grey-scale (classes are un-labeled at this juncture)

Figure 1-5 Spatial distribution and temporal trend of the $k=4$ LCA solution (years: 2001-to-2011)

1.6.4 LTA Model

Following the identification of the $k=4$ LC-model, analysis proceeded to the longitudinal setting where social fragmentation was synchronously examined across all three time points using the framework of LTA. LTA was also used to estimate the incidence of transitions in social fragmentation latent class membership from Time-1 to Time-2, and Time-2 to Time-3. As part of this process the assumption of invariance was explored by: formal tested using a nested likelihood-ratio test (i.e. G^2 -test); examining the information criteria across the nested models; and, inspecting the form and nature of the latent status profile plots from the freely estimated, non-invariant model.

With regards to the nested comparison (Table 1-5), the G^2 -test favoured a non-invariant model. However, inspection of the BIC and a-BIC values favoured measurement invariance. In addition, a visual examination of the time varying item-response profile plots from the freely estimated LTA model (Figure 1-6) did not substantively make the case for non-invariance. Considering all of this information, a decision was therefore made to impose measurement invariance across time, a facet ensured the qualitative interpretation, or meaning of the latent statuses was maintained across time aided interpretation.

Table 1-5 Results of longitudinal measurement invariance

Mx.Inv.	LL-Ho	Ho Sc.C-Fx	G^2 p-value	No. Fr.Pm.	AIC	BIC	a-BIC
No	-4613.940	1.0824		135	9497.879	10026.567	9598.255
Yes	-4688.502	1.7679	$p < 0.0001$	63	9503.004	9749.724	9549.846
<i>Mx.Inv.</i>	<i>Measurement invariance</i>						
<i>LL-Ho</i>	<i>Ho-Log-likelihood value</i>						
<i>Ho Sc.C-Fx</i>	<i>Ho scaling correction factor</i>						
<i>No. Fr.Pm.</i>	<i>Number of free parameters</i>						

With regards to the assumption of stationarity (i.e. if the propensity for neighbourhoods to transition into differing classes from Time-1 to Time-2 was analogous to that from Time-2 to Time-3), a nested model imposing stationary transition probabilities was compared to the measurement invariant model. Here again (Table 1-6) the G^2 -test and information criteria pointed in differing directions, with the former favouring the imposition of stationarity, while the later did not. And, while there distinct similarities between the two transition matrices estimated from the non-stationary model (Figure 1-7), there were also substantive probabilities on the off-diagonals, suggesting different transition patterns were likely at play across the two intervals; a factor that fed into the determination to progress with a non-stationary LTA model.

Table 1-6 Results of transition probability invariance (i.e. stationarity)

Tr. Pr.Inv.	LL-Ho	Ho ScCFx	G^2 p-value	No. Fr.Pm.	AIC	BIC	a-BIC
No	-4688.502	1.7679		63	9503.004	9749.724	9549.846
Yes	-4769.911	2.0800	$p < 0.0001$	48	9635.822	9823.800	9671.511
<i>Tr Pr.Inv.</i>	<i>Transition probability invariance</i>						
<i>LL-Ho</i>	<i>Ho-Log-likelihood value</i>						
<i>Ho Sc.C-Fx</i>	<i>Ho scaling correction factor</i>						
<i>No. Fr.Pm.</i>	<i>Number of free parameters</i>

Figure 1-6 Latent status profile plots: freely estimated (i.e. measurement non-invariance)

Proportions for the most likely membership in the four latent class statuses at each of the measurement occasions are presented in Table 1-7. And, the two transition matrices (2001-to-2006, and 2006-to-2011) for the 4-class LTA model are presented in Figure 1-7. Stasis was a feature, with most neighbourhoods remaining in the same latent status across time. Moreover, relative to classes-A and -D, neighbourhoods assigned to classes-B, and -C exhibited a higher incidence of transitioning to another class, in particular for the interval spanning 2006-to-2011. And, where the model was suggestive that neighbourhoods transitioned into another class across time, neighbourhoods predominantly shifted from class-B and -C into classes A and D.

Table 1-7 Class prevalence: $k=4$ LTA model*

	T1: 2001	T2: 2006	T3: 2011
Class-A	29.38	30.19	33.96
Class-B	17.52	17.25	15.09
Class-C	19.68	17.79	13.21
Class-D	33.42	34.77	37.74

*Prevalence based on the modal assignment of neighbourhoods to latent statuses, and measurement invariance

Figure 1-7 AR1-LTA $k=4$ transition matrices

1.7 Discussion

This study applied a multivariate finite mixture model to characterise neighbourhoods, and their evolution, with respect to social fragmentation. Four types of neighbourhoods were probabilistically identified, each defined by features reflecting distinct multidimensional combinations of attributes associated with the 'how', the 'who', and the 'why' of fragmenting social conditions. The social topography implied by these four neighbourhood types was not stagnant across the ten-year window of analysis, with a non-trivial percentage of neighbourhoods inferred to transition to another class over time. Many neighbourhoods though exhibited substantive intransience, with neighbourhoods assigned as class-A types (low social fragmentation) observed to have the greatest 'stickiness' of character.

The intransience of social fragmentation identified in this study mirrors the persistence of this trait evident in the other known study that has explored the spatial and temporal dynamics of social fragmentation at the small-area level. Grigoroglou et al. (2019) examined social fragmentation across England, and its 10 administrative regions, from 2001 to 2011, through the decennial Censuses using Congdon's (1996) four indicators (percentage of: single people, one-person households, private renting, and population turnover). The central finding of Grigoroglou et al. (2019, p.1 of 8) was the "strong persistence for social fragmentation nationally (Spearman's $r=0.93$)."

The authors also observed increased spatial clustering over time, though the reported levels of clustering remained low (Moran's $I = 0.0483$ (2001), 0.0794 (2011)).

To date, the spatial and temporal dynamics of social fragmentation have received less attention than other social traits of the environment, notably social and material deprivation (Grigoroglou et al. 2019). This study addressed this gap. More broadly, this study addressed the identified need for "longitudinal studies of neighbourhood temporal dynamics" and, the study of "how neighbourhood processes evolve over time" (Sampson et al 2002, p.472).

Study findings have a number of implications. First, these results highlight the utility of LTA as a model-based approach for classifying and categorising urban neighbourhoods with regards to traits that are consequential for social life. Neighbourhood-centred applications of LTA do not though need to be restricted to an assessment of social factors; they may also be extended to consider other multidimensional arrays of neighbourhood-level attributes, be they related to the social, built and or natural environment. Second, while, the study highlighted that the social nature of neighbourhoods can change over relatively short-to-moderate time periods, it also highlighted that many neighbourhoods do not change in character; a propensity that has been previously noted of social fragmentation (Grigoroglou et al. 2019), as well as of other traits for example neighbourhood quality of life (Delmelle and Thill 2014). However, as Galster et al. (2007, p.179) remarks, this observation does not imply "that neighbourhoods have so much inertia that they cannot be altered significantly from their stable state." Rather it suggests that time matters. From a policy orientation then the implication is that where there might be an intention to act in a targeted

manner on the social fabric of a neighbourhood “short-term, policy-induced ‘quick fixes’ hold little prospect to alter longer-term outcomes for neighbourhoods; sustained effort is required” (Galster et al 2007, p.179). Findings also underscore the importance of policy actions that account for relational dependencies among neighbourhoods. Ostensibly this inference for policy would see groups of neighbourhoods leveraged, and their collective assets brought to bear in efforts aimed at fostering social integration in focal areas.

In highlighting policy implications there is also a need to express some caution; in particular where policy may entertain ‘social mixing’ as a means to act on social fragmentation. Social mix is a nebulous concept, but one which variously relates to the socioeconomic or housing tenure profile, within a locality. Its relevance here is that social mix has often been touted as a means to foster social inclusion and build cohesive communities (Arthurson et al. 2015). However, as Arthurson et al. (2015) argues, and Lees (2008, p.2449), cogently expresses, any “policies of social mixing require critical attention with regard to their ability to produce an inclusive urban renaissance and the potentially detrimental gentrifying effects they may inflict.”

In terms of strengths, this study did not assume that neighbourhoods change in a homogenous manner. Rather neighbourhood evolution was modelled as a heterogeneous process, with neighbourhoods operationalised as multidimensional ecological units. In a similar vein, there was no a priori assumption of spatial dependence in the heterogeneity of neighbourhood-level social fragmentation. Rather, the applied latent class models allowed for the identification of dynamic profiles of neighbourhood-level social fragmentation irrespective of where they were located across the city’s spatial territory.

Further strengths relate to measurement. Latent class modelling of social fragmentation was based on a set of nine census-based indicators. Spielman and Singleton (2015, p.1007) have argued that the variance associated with a single estimate of a census-based indicator is “a symmetrically distributed random variable”. Moreover, as “variable-specific estimates are partially independent from each other, when looking at a sufficiently large collection of these variables, these random ‘errors’ will average to zero”; a feature that enables the applied multivariate approach in this study to “provide a more robust picture of the place[s] under investigation” (Spielman and Singleton 2015, p.1007).

In a related manner, Smith (2011, p.537) notes that “Chance events...are averaged out at the group level...Further, the basic notion that what is near-random at one level may be almost entirely predictable at a higher level is an emergent property of many systems, from particle physics to the social sciences”. It is a perspective that draws attention not only to the distinctiveness of the information gained by studying latent constructs, but also the utility of examining concepts at an ecological, or neighbourhood-level, as compared to a variable- or individual-level.

In considering the findings and interpretations of this study some care is required as there are a number of limitations that warrant attention; these are variously outlined and discussed below.

1.7.1 Limitations

This was a descriptive analysis of the multidimensional and developmental profile of neighbourhood-level social fragmentation. It did not set out to causally explain the estimated model. To this end a limitation is an analysis, and discussion, of the 'why' neighbourhoods evolved as they were estimated. The LTA framework is flexible and readily accommodates the capacity to model structural antecedents; both as predictors of estimated latent states at given time points, as well as predictors of transitions across time periods. Modelling of this kind is by its nature substantively more complicated and susceptible to difficulties with regards to model identification and estimation, more so where sample size is at the lower bounds of what is advised for unconditional latent transition models (an n in the range of 300-500 units).

With regards to sample size, a primary limitation relates to statistical power. Considerations relating to statistical power are very much at the forefront of methodological development in LC-applications (Baldwin 2015, Tien et al. 2013, Wurpts and Geiser 2014). Generally speaking, in LC-models there may be concerns regarding the reliability of the parameter estimates and standard errors when for example, sample size is small, and or if one or more of the latent classes exhibits an unusually low prevalence (Nylund-Gibson and Masyn 2016). For LC-models, a small sample is often expressed as an n below 300, a number that approximates the sample size in this study. However, there are many examples of meaningful neighbourhood-centred applications of LCA that have been conducted with relatively small samples. Moreover, in this study latent constructs were modelled using a moderate-to-high quality set of observed indicators, with generally good evidence of within class homogeneity and across class separation; qualities noted to enhance statistical power within latent class and transition analyses in the face of relatively small samples (Baldwin 2015, Tien et al 2013, Wurpts and Geiser 2014).

Latent class models are recognised to be dependent on the underlying sample characteristics and the observed unit-level response patterns. The corollary of these facets is that latent class and mixture models trade-off generalisability for specificity. Certainly then there needs to be recognition that although a neighbourhood-centred application of LTA can provide insight into neighbourhood-level developmental processes it remains an inductive and exploratory modelling approach.

It was also understood that the provenance of the set of latent class indicators used to measure neighbourhood social fragmentation may have impacted findings (these indicators were derived by Ivory et al. (2012) through the use of New Zealand population-level data). Concern relating to the context-specific applicability of these indicators might have been attenuated had, for example, a principal components analysis been undertaken *a priori*. Of relevance though, through a series of experiments investigating the attributes of LCA, Cole et al. (2017) demonstrated the sensitivity of the within-class item-response probabilities to variations in the measurement properties of the underlying construct of interest. To this end Cole et al. (2017, p.177) advocated that researchers applying LC-models ascertain whether their

identified class structure(s) were “replicable across minor perturbations of measurement”. Such an exercise was undertaken within this study with the indicator set added to, subtracted from, and altered.¹⁶

Another limitation was that social fragmentation was not directly measured. Ivory (2008, p.132) noted that “It would be hazardous to inflate [an indirect instrument] to be a complete measure of such a neighbourhood’s fragmenting characteristics, and even more unfortunate to assume it is a [direct] measure of the social cohesion or capital in the neighbourhood. Its use in analyses should reflect these inherent limits.” In a related manner this study did not add to the debate concerned with social selection and social causation as competing hypotheses related to neighbourhood change.

The identified typology, and set of identified transitions, are likely subject to the imposed temporal scale – two five-year intervals. Time-scale dependence in the nature of a neighbourhood’s character is consonant with the temporal aspects of the ‘uncertain geographic context problem’ articulated by Kwan (2012), and the ‘modifiable spatiotemporal unit problem’ expressed by Martin et al. (2015); itself an extension of Openshaw’s (1984) ‘modifiable areal unit problem’. Indeed, the temporal aspect of the ‘uncertain geographic context problem’ is an issue tangibly highlighted by Le Roux et al. (2017) in a Paris situated study, which over the course of a 24-hour period observed marked cyclical spatiotemporal variability in district-level social composition and social segregation; an observation that likely holds implications regarding the patterning and evolution of neighbourhood-level social fragmentation.

The temporal spacing of the observed neighbourhood-level measures at equidistant 5-year increments may also have induced temporal misclassification (Collins and Lanza 2010). The extent of such temporal misclassification would be a reflection of the speed at which neighbourhoods change relative to the temporal design features of the study (Collins 2006). However, given measurement intervals in this study were aligned with Census years it cannot be certain that neighbourhood transitions were underestimated, or indeed neighbourhood stability overestimated. Furthermore, the nature of discrete change profiles generated by FMM, such as LTA, may be sensitive to the number of encompassed time-points (Jackson and Sher 2006). Though as Collins and Lanza (2010) note, the interpretation of transition probabilities should always be relative to the model, and any inferences beyond this should be made with caution.

There was a discrepancy between the boundaries of Gazetted Localities (SSCs) and ABS supported SSCs. For example, in 2011 SSCs were derived on the basis of whole SA1s whereas the official gazetted suburban boundary applied different criteria (variously land parcel boundaries, natural geographic cleavages etc.). Moreover, with the genesis of the Australian Statistical Geography Standard (ASGS) (a classification which

¹⁶ In preliminary analyses (not reported here), LCA models were variously estimated using two versions of the Language spoken indicator: “Main language spoken at Home if not English” and “Proficiency of English language”. In addition, models were estimated with a ‘Social Housing’ Indicator in two different flavours, with and without the ‘Home Ownership’ indicator in the model. Furthermore, models were estimated with and without either of the ‘Same Usual Residence x1 / x5-year’ indicators in the model. Analyses available on request.

supplanted the Australian Standard Geographical Classification (ASGC)), SA1s replaced CDs as the base unit geographies. Arguably, these boundary and methodological changes may have induced non-differential misclassification bias (Blakely and Woodward 2000). However, it is unclear what influence non-differential misclassification bias would have had on the formation of the latent class models.

Neighbourhoods included in these analyses were operationalised through the SSC. This specification (and restriction) brings to the fore the spatial aspects of the aforementioned modifiable areal unit (Openshaw 1984), as well as uncertain geographic context problem (Kwan 2012) and, their respective implications for inference. For example, Browning et al. (2017, p.227) observed that the pattern of social integration and segregation may co-vary not only on the basis of residential factors, but may conditionally extend to the “activity locations neighbourhood residents frequent in the course of their daily routines.”

At the heart of this study is classification and categorisation. Curran and Bauer (2016, para.1) point toward a tension where “Extracting categories when variation is really continuous presents risks, but so too does failing to identify meaningfully distinct subgroups”. Categories may quieten noisy information, and order a messy world. But as Vanderbilt (2016, para.7) notes, this “categorical perception...is not an innocent process”. Moreover, Vanderbilt (2016, para.8) argues that although “Similarity serves as a basis for the classification of objects”, similarities may be perceptual artefacts of the imposed classificatory system. And, relatedly “Things we might have viewed as more similar become, when placed into distinct categories, more different.” For Vanderbilt (2016, para.22), the corollary of this dependence “on categorising, is that we could miss something outside our perception”. And, that while grappling with complex information may induce cognitive disfluency (Owen et al. 2016), it might function to prompt deeper and more thoughtful engagement (Alter 2013) which might be nullified through the simplifying effect of categorisation. An offset to these perspectives is that in the context of this study the latent typology was identified through an inductive model-based process.

This study also applied median splits to the latent class indicators. The case against median splits is well-known, particularly for applications that might otherwise see continuous independent variables analysed using linear models. Central arguments related to median splits include the loss of information concerning intra-subject differences, loss of statistical power, and the induction of spurious statistical relationships (MacCallum et al. 2002). Indeed the totality of these issues led MacCallum et al. (2002, p.19) to state that “dichotomization is rarely defensible and often will yield misleading results”. However, in a series of papers, based on a simulation analysis, Iacobucci et al. (2015a, 2015b) defended the use of median splits, in particular for analyses where a set of independent variables are uncorrelated; a condition that features as one of the fundamental assumptions of latent class models, and which was an assumption that was met in this study. There are also recent findings from LC-specific simulation analyses, albeit involving latent classes reflective of count-based indicators, that median split dichotomisation may not affect the quality of mixture

recovery, or bias associated parameter estimates, particularly in circumstances where the number of estimated classes exceeds the number of indicators (Macia and Wickham 2018).

Putting aside the respective arguments of Iacobucci et al. (2015b), Macia and Wickham (2018), and MacCallum et al. (2002), what of the categorical alternatives to applying median splits in this study? For example, Gelman and Park (2008) advanced a case for trichotomising variables. However, trichotomising nine indicators in an LTA, measured across three-time points, would have seen the possible response pattern expand from 2^{27} to 3^{27} , a likely unsustainable increase.

Finally, neighbourhoods were assigned to latent statuses based on their highest posterior probabilities of membership; a factor that did not account for the uncertainty in latent membership. The issue of uncertainty and classification error in both cross-sectional and longitudinal variants of latent class models is a function of the identification of a well-defined and meaningful model that innately suppresses classification error and assignment uncertainty. Moreover, given the high classification quality of the LCA and LTA models (entropy exceeding 0.90), classification error that may have been introduced could be surmised to have been small. In addition, analyses presented in this paper were descriptive in nature.

1.7.2 Future research

By drawing on, and extending the theoretical and methodological approach adopted in this study, there are a number of areas of enquiry that might be pursued to strengthen understandings of neighbourhood-level social fragmentation; a number of these are outlined below.

Arguably, cities and their neighbourhoods are products of historic and ongoing socio-political processes. Or, as Piiparinen (2017, para.15) remarks “It’s on the inside: in our fears, and inside the systems that simultaneously placate and preen them” that “Power and politics make places, not the other way around.” To this end, if neighbourhoods are both products of power and politics, as well as producers of social patterns, then there is need to more actively engage with these factors. A fuller understanding of the manner by which socio-political processes shape the topography and dynamics of social fragmentation might be gained by drawing on critical theory to complement analytic enquiries. As a critical theory, political ecology, in particular, is germane to the study of neighbourhoods, and neighbourhood evolution, as it addresses the impacts of structural factors such as power, race, and class (Chitewere et al. 2017). Political ecology has its roots in political economy (the examination of the unequal distribution of power and wealth in a society), and cultural ecology (the analysis of people, their way of life and their physical environment) (Chitewere et al. 2017). Political ecology is also tied to the theories of environmental and social justice. With equity in mind, using political ecology to question *why* neighbourhood types are, and *how* they came to be and the manner by which they morph, might unlock insights into the non-random distribution of neighbourhood types, the manner in which neighbourhood types are differentially

experienced by populations (*for whom*), and the likely mode of action requisite to addressing the just nature of the social conditions of living.

As expressed, a lifecourse perspective affords a framework to study and understand neighbourhoods and neighbourhood dynamics. Fundamental factors in the lifecourse approach are: age, time (or period), and cohort year. Conjointly bringing these three factors to bear in an analysis would offer a unique way of examining competing, intersecting and or coalescent influences on the evolution of social contexts. For example, in a neighbourhood-centred typological analysis, age effects might be viewed as those tied to a neighbourhood's chronological development; while, period effects might reflect time-limited macro-social influences (such as fluxes in economic and housing market circumstances); with, cohort effects construed to reflect neighbourhood-traits such as the prevailing design-type or spatial syntax. Relational factors are also a feature of a lifecourse-of-place perspective. To that end the flexibility of the LTA framework could be applied to model and examine the effects of relational antecedents. For example, modelling the effects of adjacent neighbourhoods' social fragmentation status on the development of focal areas. Or, considering how a neighbourhood's social fragmentation status changes over time as a function of its structural relationship to social services and resources valued by its residents (Livingston et al. 2013).

A third research direction to strengthen understanding of neighbourhood dynamics is to investigate the interaction of exogenous factors with endogenous processes. As Galster et al. (2007) noted, beyond the classic work of Thomas Schelling, an economist who examined the spatial dynamics of race, there has been limited empirical analysis of a neighbourhood's internal processes that might see it shift in one direction or another after it has been perturbed from equilibrium or, that function as self-correcting mechanisms to stabilise it in the face of a shock. A promising approach to the study of endogenous stability at the neighbourhood-level is to integrate methods such as LTA with a complex systems model such as cellular automata. Indeed, there is an applied history of integrating Markov models and Cellular Automata, though most extant applications have focused on land use (Ghosh et al. 2017). Simulation models such as cellular automata additionally provide a platform for synthetic experimentation, a functionality that affords a test bed for examining a range of thought experiments, mechanisms and policy interventions.

1.8 Conclusion

In conclusion, in this spatio-temporal analysis four neighbourhood types were identified that variously reflected diverse dimensional profiles of social fragmentation, giving rise to a variegated metropolitan social topography. While certain neighbourhoods were inferred to have changed type across time, with modelled shifts in the nature of their fragmenting profile, most neighbourhoods were characterised by intransience. And, although analyses are exploratory and descriptive, collectively the insights into neighbourhood change and the developmental evolution of social fragmentation may be relevant for framing actions that aim to impact the social fabric at the neighbourhood level.

1.9 References

- Alter, A.L., 2013. The benefits of cognitive disfluency. *Current Directions in Psychological Science* 22, 437-442.
- Apparicio, P., Riva, M., Séguin, A.-M., 2015. A comparison of two methods for classifying trajectories: A case study on neighbourhood poverty at the intra-metropolitan level in Montreal. *Cybergeog: European Journal of Geography*, doi: 10.4000/cybergeog.27035.
- Arthurson, K., Levin, I., Ziersch, A., 2015. What is the meaning of 'social mix'? Shifting perspectives in planning and implementing public housing estate redevelopment. *Australian Geographer* 46, 491-505.
- Australian Bureau of Statistics, 2001a. Census of Population and Housing: Census Geographic Areas Digital Boundaries: Catalogue No. 2923.0.30.001 Australian Bureau of Statistics, Canberra, Australia.
- Australian Bureau of Statistics, 2001b. Census of Population and Housing: Basic Community Profile, 2001 First Release: Catalogue No. 2001.0. Australian Bureau of Statistics, Canberra, Australia.
- Australian Bureau of Statistics, 2006a. Census of Population and Housing: Basic Community Profile, 2006 First Release: Catalogue No. 2001.0. Australian Bureau of Statistics, Canberra, Australia.
- Australian Bureau of Statistics, 2006b. Statistical Geography - Australian Standard Geographical Classification (ASGC), Digital Boundaries, 2006: Catalogue No. 1259.0.30.002 Australian Bureau of Statistics, Canberra, Australia.
- Australian Bureau of Statistics, 2011a. Australian Social Trends, December 2011. Catalogue no. 4102.0. Australian Bureau of Statistics, Canberra, Australia.
- Australian Bureau of Statistics, 2011b. Australian Standard Geographical Classification (ASGC), Digital Boundaries, 2011: Catalogue No. 1259.0.30.001 Australian Bureau of Statistics, Canberra, Australia.
- Australian Bureau of Statistics, 2011c. Census of Population and Housing: Basic Community Profile, 2011 First Release: Catalogue No. 2001.0. Australian Bureau of Statistics, Canberra, Australia.
- Australian Bureau of Statistics, 2012. Statistical Geography: Statistical Geography Fact Sheet - Greater Capital City Statistical Areas. Australian Bureau of Statistics, Canberra, Australia.
- Australian Bureau of Statistics, 2016. Australian Statistical Geography Standard (ASGS): Volume 3 - Non ABS Structures, July 2016: Catalogue No. 1270.0.55.003. Australian Bureau of Statistics, Canberra, Australia.
- Bakk, Z., Oberski, D.L., Vermunt, J.K., 2014. Relating latent class assignments to external variables: Standard errors for correct inference. *Political Analysis* 22, 520-540.
- Baldwin, E.E., 2015. A Monte Carlo Simulation Study Examining Statistical Power in Latent Transition Analysis. University of California, Santa Barbara, Santa Barbara, Calif.
- Baum, F., Palmer, C., 2002. 'Opportunity structures': Urban landscape, social capital and health promotion in Australia. *Health Promotion International* 17, 351-361.
- Bell, A., Jones, K., 2015. Explaining fixed effects: Random effects modelling of time-series cross-sectional and panel data. *Political Science Research and Methods* 3, 133-153.
- Berkman, L.F., Glass, T., Brissette, I., Seeman, T.E., 2000. From social integration to health: Durkheim in the new millennium. *Social Science and Medicine* 51, 843-857.
- Berlin, K.S., Williams, N.A., Parra, G.R., 2014. An introduction to latent variable mixture modeling (Part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology* 39, 174-187.
- Blakely, T.A., Woodward, A.J., 2000. Ecological effects in multi-level studies. *Journal of Epidemiology and Community Health* 54, 367-374.
- Bolck, A., Croon, M., Hagenaars, J., 2004. Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis* 12, 3-27.

- Browning, C.R., Calder, C.A., Krivo, L.J., Smith, A.L., Boettner, B., 2017. Socioeconomic segregation of activity spaces in urban neighborhoods: Does shared residence mean shared routines? *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3, 210-231.
- Bruch, E.E., Mare, R.D., 2006. Neighborhood choice and neighborhood change. *American Journal of Sociology* 112, 667-709.
- Building Healthy Communities in California, 2017.
- Celeux, G., Soromenho, G., 1996. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification* 13, 195-212.
- Chitewere, T., Shim, J.K., Barker, J.C., Yen, I.H., 2017. How neighborhoods influence health: Lessons to be learned from the application of political ecology. *Health & Place* 45, 117-123.
- Clark, T.N., 2008. Program for a new Chicago School. *Urban Geography* 29, 154-166.
- Coffee, N.T., Lange, J., Baker, E., 2016. Visualising 30 years of population density change in Australia's major capital cities. *Australian Geographer* 47, 511-525.
- Cole, V.T., Bauer, D.J., Hussong, A.M., Giordano, M.L., 2017. An empirical assessment of the sensitivity of mixture models to changes in measurement. *Structural Equation Modeling: A Multidisciplinary Journal* 24, 159-179.
- Collins, L.M., 2006. Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology* 57, 505-528.
- Collins, L.M., Lanza, S.T., 2010. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. John Wiley & Sons, Inc., Hoboken, NJ.
- Congdon, P., 1996. Suicide and parasuicide in London: A small-area study. *Urban Studies* 33, 137-158.
- Congdon, P., 2004. Commentary: Contextual effects: Index construction and technique. *International Journal of Epidemiology* 33, 741-742.
- Curran, P.J., Bauer, D.J., 2016. The Human Desire to Categorize, Curran-Bauer Analytics.
- Davison, G., 1995. Australia: The first suburban nation? *Journal of Urban History* 22, 40-74.
- Delmelle, E., 2015. Five decades of neighborhood classifications and their transitions: A comparison of four cities, 1970-2010. *Applied Geography* 57, 1-11.
- Delmelle, E., Thill, J.-C., 2014. Neighborhood quality-of-life dynamics and the Great Recession: The case of Charlotte, North Carolina. *Environment and Planning A* 46, 867-884.
- Delmelle, E., Thill, J.-C., Furuseth, O., Ludden, T., 2013. Trajectories of multidimensional neighbourhood quality of life change. *Urban Studies* 50, 923-941.
- Delmelle, E., Thill, J.-C., Wang, C., 2016. Spatial dynamics of urban neighborhood quality of life. *The Annals of Regional Science* 56, 687-705.
- Delmelle, E.C., 2016. Mapping the DNA of urban neighborhoods: Clustering longitudinal sequences of neighborhood socioeconomic change. *Annals of the American Association of Geographers* 106, 36-56.
- Dovey, K., Ristic, M., 2017. Mapping urban assemblages: The production of spatial knowledge. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* 10, 15-28.
- Dziak, J.J., Coffman, D.L., Lanza, S.T., Li, R., Jermiin, L.S., 2019. Sensitivity and specificity of information criteria. *bioRxiv* 449751; doi: <https://doi.org/10.1101/449751>.
- Enders, C.K., Gottschall, A.C., 2011. Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal* 18, 35-54.
- ESRI, 2009. GIS best practices: social sciences, in: ESRI (Ed.). ESRI.
- Fagg, J., Curtis, S., Stansfeld, S.A., Cattell, V., Tupuola, A.-M., Arephin, M., 2008. Area social fragmentation, social support for individuals and psychosocial health in young adults: Evidence from a national survey in England. *Social Science and Medicine* 66, 242-254.

- Florida, R., 2017. Two Takes on the Fate of Future Cities. CityLab, 21 April 2017 (accessed 23 May 2017): <https://www.citylab.com/equity/2017/04/two-takes-on-the-fate-of-future-cities/521907/>.
- Frankfurt, S., Frazier, P., Syed, M., Jung, K.R., 2016. Using group-based trajectory and growth mixture modeling to identify classes of change trajectories. *The Counseling Psychologist* 44, 622-660.
- Friel, S., Akerman, M., Hancock, T., Kumaresan, J., Marmot, M., Melin, T., Vlahov, D., members, G., 2011. Addressing the social and environmental determinants of urban health equity: Evidence for action and a research agenda. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 88, 860-874.
- Galster, G., 2001. On the nature of neighbourhood. *Urban Studies* 38, 2111-2124.
- Galster, G., Cutsinger, J., Lim, U., 2007. Are neighbourhoods self-stabilising? Exploring endogenous dynamics. *Urban Studies* 44, 167-185.
- Geiser, C., 2013. *Data analysis with Mplus (English Ed.)*. The Guilford Press, New York, NY.
- Gelman, A., Park, D.K., 2009. Splitting a predictor at the upper quarter or third and the lower quarter or third. *The American Statistician* 63, 1-8.
- Ghosh, P., Mukhopadhyay, A., Chanda, A., Mondal, P., Akhand, A., Mukherjee, S., Nayak, S.K., Ghosh, S., Mitra, D., Ghosh, T., Hazra, S., 2017. Application of cellular automata and Markov-chain model in geospatial environmental modeling- a review. *Remote Sensing Applications: Society and Environment* 5, 64-77.
- Goodchild, M.F., Lam, N.S.-N., 1980. Areal interpolation: A variant of the traditional spatial problem. *Geo-processing* 1, 297-312.
- Gregory, I.N., Ell, P.S., 2005. Breaking the boundaries: Geographical approaches to integrating 200 years of the census. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168, 419-437.
- Grigoroglou, C., Munford, L., Webb, R.T., Kapur, N., Ashcroft, D.M. and Kontopantelis, E., 2019. Spatial distribution and temporal trends in social fragmentation in England, 2001– 2011: a national study. *BMJ Open*, 9(1), p.e025881.
- Gunasekara, F.I., Richardson, K., Carter, K., Blakely, T., 2014. Fixed effects analysis of repeated measures data. *International Journal of Epidemiology* 43, 264-269.
- Hanibuchi, T., Kondo, K., Nakaya, T., Shirai, K., Hirai, H., Kawachi, I., 2012a. Does walkable mean sociable? Neighborhood determinants of social capital among older adults in Japan. *Health & Place* 18, 229-239.
- Hanibuchi, T., Murata, Y., Ichida, Y., Hirai, H., Kawachi, I., Kondo, K., 2012b. Place-specific constructs of social capital and their possible associations to health: a Japanese case study. *Social Science and Medicine* 25, 225-232.
- Harris, R., Feng, Y., 2016. Putting the geography into geodemographics: Using multilevel modelling to improve neighbourhood targeting – a case study of Asian pupils in London. *Journal of Marketing Analytics* 4, 93-107.
- Hipp, J.R., Bauer, D.J., 2006. Local solutions in the estimation of growth mixture models. *Psychological Methods* 11, 36-53.
- Hirsch, J.A., Green, G.F., Peterson, M., Rodriguez, D.A., Gordon-Larsen, P., 2017. Neighborhood sociodemographics and change in built infrastructure. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* 10, 181-197.
- Howard, N.J., 2011. *Whose place is it? Examining the socio-spatial geography of obesity in young adults for an Australian context*, Discipline of Geographical and Environmental Studies, Faculty of Humanities and Social Sciences. University of Adelaide, South Australia, Adelaide, p. 297pp.
- Iacobucci, D., Posavac, S.S., Kardes, F.R., Schneider, M.J., Popovich, D.L., 2015a. The median split: Robust, refined, and revived. *Journal of Consumer Psychology* 25, 690-704.
- Iacobucci, D., Posavac, S.S., Kardes, F.R., Schneider, M.J., Popovich, D.L., 2015b. Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology* 25, 652-665.
- Ivory, V.C., 2008. 'Neighbourhood Social Fragmentation and Health. Bringing Social Epidemiology and Social Theory Together', PhD Thesis. University of Otago, Dunedin, New Zealand, Otago, NZ, p. 181pp.

- Ivory, V., Witten, K., Salmond, C., Lin, E.Y., Quan, Y.R., Blakely, T., 2012. The New Zealand Index of Neighbourhood Social Fragmentation: Integrating theory and data. *Environment and Planning A* 44, 972-988.
- Ivory, V.C., Collings, S.C., Blakely, T., Dew, K., 2011. When does neighbourhood matter? Multilevel relationships between neighbourhood social fragmentation and mental health. *Social Science and Medicine* 72, 1993-2002.
- Jackson, K.M., Sher, K.J., 2006. Comparison of longitudinal phenotypes based on number and timing of assessments: A systematic comparison of trajectory approaches II. *Psychology of Addictive Behaviors* 20, 373-384.
- Jones, R., Heim, D., Hunter, S., Ellaway, A., 2014. The relative influence of neighbourhood incivilities, cognitive social capital, club membership and individual characteristics on positive mental health. *Health & Place* 28, 187-193.
- Joseph, L., 2008. Finding space beyond variables: An analytical review of urban space and social inequalities. *Spaces for Difference: An Interdisciplinary Journal* 1, 29-50.
- Kim, M., Vermunt, J., Bakk, Z., Jaki, T., Van Horn, M.L., 2016. Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal* 23, 601-614.
- Kline, R., 2011. *Principles and Practice of Structural Equation Modeling* (2nd ed.). New York: The Guilford Press.
- Kupke, V., Rossini, P., McGreal, S., 2011. A multivariate study of medium density housing development and neighbourhood change within Australian cities. *Pacific Rim Property Research Journal* 17, 3-23.
- Kwan, M.-P., 2012. The uncertain geographic context problem. *Annals of the Association of American Geographers* 102, 958-968.
- Lanza, S.T., Bray, B.C., 2010. Transitions in drug use among high-risk women: An application of latent class and latent transition analysis. *Advances and Applications in Statistical Sciences* 3, 203-235.
- Lawless, P., Foden, M., Wilson, I., Beatty, C., 2010. Understanding area-based regeneration: The New Deal for Communities Programme in England. *Urban Studies* 47, 257-275.
- Le Roux, G., Vallée, J., Commenges, H., 2017. Social segregation around the clock in the Paris region (France). *Journal of Transport Geography* 59, 134-145.
- Lees, L., 2008. Gentrification and social mixing: Towards an inclusive urban renaissance? *Urban Studies* 45, 2449-2470.
- Lekkas, P., Paquet, C., Howard, N.J., Daniel, M., 2017a. Illuminating the lifecourse of place in the longitudinal study of neighbourhoods and health. *Social Science and Medicine* 177, 239-247.
- Lekkas, P., Paquet, C., Howard, N.J., Daniel, M., 2017b. The lifecourse of place: Looking past paradigms and metaphors to the just nature of place-health – A rejoinder to Andrews'. *Social Science and Medicine* 175, 215-218.
- Ling, C., Delmelle, E.C., 2016. Classifying multidimensional trajectories of neighbourhood change: A self-organizing map and k-means approach. *Annals of GIS*, 1-14.
- Livingston, M., Walsh, D., Whyte, B., Bailey, N., 2013. Investigating the Impact of the Spatial Distribution of Deprivation on Health Outcomes. Project Report. Glasgow Centre for Population Health, University of Scotland, Glasgow, UK.
- Logan, J.R., Xu, Z., Stults, B.J., 2014. Interpolating U.S. decennial census tract data from as early as 1970 to 2010: A longitudinal tract database. *The Professional Geographer* 66, 412-420.
- Lubke, G., Neale, M.C., 2006. Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research* 41, 499-532.
- Lucumí, D.I., Gomez, L.F., Brownson, R.C., Parra, D.C., 2015. Social capital, socioeconomic status, and health-related quality of life among older adults in Bogotá (Colombia). *Journal of Aging and Health* 27, 730-750.

- Lutters, W.G., Ackerman, M.S., 1996. An introduction to the Chicago School of Sociology, Interval Research Proprietary.
- MacCallum, R.C., Zhang, S., Preacher, K.J., Rucker, D.D., 2002. On the practice of dichotomization of quantitative variables. *Psychological Methods* 7, 19-40.
- Macia KS, Wickham RE. 2018. The impact of item misspecification and dichotomization on class and parameter recovery in LCA of count data. *Multivariate Behavioral Research*.
<https://doi.org/10.1080/00273171.2018.1499499>, 33pp.
- Mair, C., Diez Roux, A.V., Golden, S.H., Rapp, S., Seeman, T., Shea, S., 2015. Change in neighborhood environments and depressive symptoms in New York City: The Multi-Ethnic Study of Atherosclerosis. *Health & Place* 32, 93-98.
- Martin, D., Cockings, S., Leung, S., 2015. Developing a flexible framework for spatiotemporal population modeling. *Annals of the Association of American Geographers* 105, 754-772.
- Masyn, K.E., 2013. Latent Class Analysis and Finite Mixture Modeling, in: Little, T.D. (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology*. Oxford University Press, New York, pp. 551-611.
- McDonald, K., Hearst, M., Farbaksh, K., Patnode, C., Forsyth, A., Sirard, J., Lytle, L., 2012. Adolescent physical activity and the built environment: A latent class analysis approach. *Health & Place* 18, 191-198.
- Meen, G., Nygaard, C., Meen, J., 2013. The Causes of Long-Term Neighbourhood Change, in: van Ham, M., Manley, D., Bailey, N., Simpson, L., Maclennan, D. (Eds.), *Understanding Neighbourhood Dynamics: New Insights for Neighbourhood Effects Research*. Springer Netherlands, pp. 43-62.
- Meyer, K.A., Boone-Heinonen, J., Duffey, K.J., Rodriguez, D.A., Kiefe, C.I., Lewis, C.E., Gordon-Larsen, P., 2015. Combined measure of neighborhood food and physical activity environments and weight-related outcomes: The CARDIA study. *Health & Place* 33, 9-18.
- Mikelbank, B.A., 2011. Neighborhood déjà vu: Classification in metropolitan Cleveland, 1970-2000. *Urban Geography* 32, 317-333.
- Morenoff, J.D., Tienda, M., 1997. Underclass neighborhoods in temporal and ecological perspective. *Annals of the American Academy of Political and Social Science* 551, 59-72.
- Moya, A.R.Z., Yáñez, C.J.N., 2017. Impact of area regeneration policies: Performing integral interventions, changing opportunity structures and reducing health inequalities. *Journal of Epidemiology and Community Health* 71, 239-247.
- Muthén, L.K., Muthén, B.O., 1998-2017. *Mplus user's guide*. Eighth Edition. Muthén & Muthén, Los Angeles, Ca.
- Muthén B, 2009. "Diverging LMR and BLRT p-values", Mplus Discussion Board,
<http://www.statmodel.com/discussion/messages/13/4529.html?1249068377> Accessed June 25th 2018.
- Nagin, D., 2005. *Group-based modeling of development*. Harvard University Press, Cambridge, MA.
- Norman, P., 2004. *Constructing a sociodemographic data time-series: Computational issues and solutions*, ESRC Research Methods Programme. University of Manchester, Manchester.
- Norman, P., Rees, P., Boyle, P., 2003. Achieving data compatibility over space and time: Creating consistent geographical zones. *International Journal of Population Geography* 9, 365-386.
- Nylund-Gibson, K., Masyn, K.E., 2016. Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal* 23, 782-797.
- Nylund, K.L., 2007. *Latent Transition Analysis: Modeling Extensions and an Application to Peer Victimization*. University of California, Los Angeles, Ca. p. 190.
- Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling* 14, 535-569.
- Openshaw, S., 1984. Ecological fallacies and the analysis of areal census data. *Environment and Planning A* 16, 17-31.

- Owen, H.E., Halberstadt, J., Carr, E.W., Winkielman, P., 2016. Johnny Depp, reconsidered: How category-relative processing fluency determines the appeal of gender ambiguity. *PloS One* 11, e0146328.
- Owens, A., 2012. Neighborhoods on the rise: A typology of neighborhoods experiencing socioeconomic ascent. *City & Community* 11, 345-369.
- Palumbo, A., Michael, Y., Hyslop, T., 2016. Latent class model characterization of neighborhood socioeconomic status. *Cancer Causes and Control* 27, 445-452.
- Piiparinen, R., 2017. The Cracks in the Melting Pot. CityLab, 8 Nov 2017 (accessed 21 Nov 2017): <https://www.citylab.com/equity/2017/11/the-cracks-in-the-melting-pot/545309/>.
- Randolph, B., Freestone, R., 2008. Problems and prospects for suburban renewal: an Australian perspective, 1st ed. City Futures Research Centre, Faculty of the Built Environment, University of New South Wales, Sydney.
- Reboussin, B.A., Ip, E.H., Wolfson, M., 2008. Locally dependent latent class models with covariates: An application to under-age drinking in the USA. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171, 877-897.
- Richardson, A.S., Meyer, K.A., Howard, A.G., Boone-Heinonen, J., Popkin, B.M., Evenson, K.R., Kiefe, C.I., Lewis, C.E., Gordon-Larsen, P., 2014. Neighborhood socioeconomic status and food environment: A 20-year longitudinal latent class analysis among CARDIA participants. *Health & Place* 30, 145-153.
- Salvati, L., Carlucci, M., Serra, P., 2018. Unraveling latent dimensions of the urban mosaic: A multi-criteria spatial approach to metropolitan transformations. *Environment and Planning A: Economy and Space*, 50, 93-110.
- Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., 2002. Assessing "neighborhood effects": social processes and new directions in research. *Annual Review of Sociology* 28, 443-478.
- Sampson, R., Morenoff, J.D., 2006. Durable inequality: Spatial dynamics, social processes and the persistence of poverty in Chicago neighborhoods, in: Bowles, S., Durlauf, S.N., Hoff, K. (Eds.), *Poverty Traps*. Princeton University Press, Princeton, NJ, pp. 176-203.
- Sampson, R., Sharkey, P., 2008. Neighborhood selection and the social reproduction of concentrated racial inequality. *Demography* 45, 1-29.
- Sampson, R.J., Schachner, J.N., Mare, R.D., 2017. Urban income inequality and the Great Recession in sunbelt form: Disentangling individual and neighborhood-level change in Los Angeles. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3, 102-128.
- Schmidt, N.M., Tchetgen Tchetgen, E.J., Ehntholt, A., Almeida, J., Nguyen, Q.C., Molnar, B.E., Azrael, D., Osypuk, T.L., 2014. Does neighborhood collective efficacy for families change over time? The Boston Neighborhood Survey. *Journal of Community Psychology* 42, 61-79.
- Schwirian, K.P., 1983. Models of neighborhood change. *Annual Review of Sociology* 9, 83-102.
- Séguin, A.-M., Apparicio, P., Riva, M., Negron-Poblete, P., 2015. The changing spatial distribution of Montreal seniors at the neighbourhood level: A trajectory analysis. *Housing Studies* 31, 61-80.
- Simic, Z., 2008. 'What are ya?': Negotiating identities in the western suburbs of Sydney during the 1980s. *Journal of Australian Studies* 32, 223-236.
- Simpson, L., 2002. Geography conversion tables: A framework for conversion of data between geographical units. *International Journal of Population Geography* 8, 69-82.
- Smith, G.D., 2011. Epidemiology, epigenetics and the 'Gloomy Prospect': Embracing randomness in population health research and practice. *International Journal of Epidemiology* 40, 537-562.
- Solari, C.D., 2012. Affluent neighborhood persistence and change in U.S. cities. *City & Community* 11, 370-388.
- Spielman, S.E., Singleton, A., 2015. Studying neighborhoods using uncertain data from the American Community Survey: A contextual approach. *Annals of the Association of American Geographers* 105, 1003-1025.

- Tein, J.-Y., Coxe, S., Cham, H., 2013. Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal* 20, 640-657.
- United States Census Bureau, 2001. Census 2000 Summary File 1 Technical Documentation: Appendix A. Census 2000 Geographic Terms and Concepts [online webpage]. Available at: <http://www.census.gov/prod/cen2000/doc/sf1.pdf> [Accessed July 10th 2016].
- Vanderbilt, T., 2016. The psychology of genre: Why we don't like what we struggle to categorize, *New York Times*, 28 May, (viewed 24 June 2017): https://www.nytimes.com/2016/05/29/opinion/sunday/the-psychology-of-genre.html?_r=0.
- Vermunt, J.K., 2010. Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis* 18, 450-469.
- Wall, M.M., Larson, N.I., Forsyth, A., Van Riper, D.C., Graham, D.J., Story, M.T., Neumark-Sztainer, D., 2012. Patterns of obesogenic neighborhood features and adolescent weight: A comparison of statistical approaches. *American Journal of Preventive Medicine* 42, e65-e75.
- Warner, T.D., Settersten Jr, R.A., 2016. Why Neighborhoods (and How We Study Them) Matter for Adolescent Development, *Advances in Child Development and Behavior*. JAI.
- Weden, M.M., Bird, C.E., Escarce, J.J., Lurie, N., 2010. Technical detail and appendices for a study of neighborhood archetypes for population health research. RAND Corporation, Santa Monica, CA, p. 50pp.
- Weden, M.M., Bird, C.E., Escarce, J.J., Lurie, N., 2011. Neighborhood archetypes for population health research: Is there no place like home? *Health & Place* 17, 289-299.
- Wei, F., Knox, P.L., 2014. Neighborhood change in metropolitan America, 1990 to 2010. *Urban Affairs Review* 50, 459-489.
- White, J., Greene, G., Farewell, D., Dunstan, F., Rodgers, S., Lyons, R.A., Humphreys, I., John, A., Webster, C., Phillips, C.J., Fone, D., 2017. Improving mental health through the regeneration of deprived neighborhoods: A natural experiment. *American Journal of Epidemiology* 186, 473-480.
- Wight, R.G., Aneshensel, C.S., Barrett, C., Ko, M., Chodosh, J., Karlamangla, A.S., 2013. Urban neighbourhood unemployment history and depressive symptoms over time among late middle age and older adults. *Journal of Epidemiology and Community Health* 67, 153-158.
- World Health Organization, 2016. Glossary of terms used, Health Impact Assessment (HIA), World Health Organization, viewed 13 June 2016. European Centre for Health Policy, WHO Regional Office for Europe, Brussels.
- Wurpts, I.C., Geiser, C., 2014. Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in Psychology* 5 (August), 1-15.

Author contributions

Peter Lekkas initiated the idea for the research, conceptualised the design and analysis of this study, extracted and derived all the variables in the development of the spatiotemporal dataset as well as in the analyses, conducted all analyses, interpreted the results, wrote the first draft of the paper, and critically revised all subsequent versions of the paper.

Natasha J Howard assisted to refine the study design, oversaw the development of the spatiotemporal dataset, provided critical feedback, and suggested revisions to the paper.

Ivana Stankov assisted to refine the study design, critically discussed the development of the spatiotemporal database, assisted in the interpretation of the study findings, provided critical feedback, and suggested revisions to the paper.

Mark Daniel discussed the study design and provided critical feedback and suggested revisions to the paper.

Catherine Paquet assisted to refine the study design, assisted in the interpretation of the study findings, provided critical feedback, and suggested revisions to the paper.

Funding and disclosures

Mr Peter Lekkas was supported by: a Research Training Scheme award (formerly the Australian Postgraduate Award), Department of Education and Training, Australian Government; a University of South Australia Scholarship; and the School of Health Sciences, University of South Australia through the Research Chair, Social Epidemiology.

Dr Catherine Paquet was funded by a National Health and Medical Research Council (Australia) Program Grant (#0631947).

Adjunct Prof. Mark Daniel was funded by a Research Chair, Social Epidemiology, University of South Australia.

Acknowledgements

The authors are grateful for the discussion and constructive feedback provided by Dr Julie Collins, at the University of South Australia.

Appendix A LTA: Extended methods

Mathematical model

Mathematically, the LTA model is expressed as per equation-C1:¹⁷

$$prob(Y = y) = \sum_{p=1}^S \sum_{q=1}^S \delta_p \rho_{i'|p} \rho_{j'|p} \rho_{k'|p} \rho_{i''|p} \rho_{j''|p} \rho_{k''|p} \tau_{q|p} \quad [C1]$$

where, Y equals the proportion of units generating a particular response y , and $t = 1^{st}$ time of measurement, $t + 1 = 2^{nd}$ time of measurement, $i' - i'' =$ response categories 1, 2... I for 1^{st} indicator, $j' - j'' =$ response categories 1, 2... J for 2^{nd} indicator, $k' - k'' =$ response categories 1, 2... K for 3^{rd} indicator, $i', j', k' =$ responses obtained at time t , $i'', j'', k'' =$ responses obtained at time $t + 1$, $p =$ latent status at time t and $q =$ latent status at time $t + 1$.

Estimation process

LTA is undertaken with **maximum likelihood estimation**. During this process, for each model run, parameter values are estimated through an iterative process using **maximum likelihood (ML) estimation** with an **expectation-maximization (EM) algorithm** (Do and Batzoglou 2008). The expectation-maximization algorithm iterates between an 'E-step', that aims at "guessing a probability distribution over completions of missing data [the latent or hidden variable] given the current model", and an 'M-step' "re-estimating the model parameters using these completions" to maximise the expected log-likelihood of the data found in the 'E-step' (Do and Batzoglou 2008, p.898). However, difficulties with likelihood surfaces can hamper the **identification of latent class models** (Masyn 2013). These difficulties, encompassing multiple local maxima and flattish saddle points, complicate the identification of a global maximum likelihood (ML) solution. Models with few parameter restrictions can be more prone to likelihood functions characterised by multiple modes. The tenability of model identification will therefore be a consideration when attention turns to specifying the form of measurement invariance in the LTA model.

Model identification can though be assisted by applying multiple sets of random starting values for the iterative estimation algorithm (Hipp and Bauer 2006); this permitting a denser search of the likelihood parameter space and an enhanced prospect of identifying the **global maxima** (highest likelihood).

Recommendations are to specify a minimum of 100-to-500 sets of random start values for complex models (Geiser 2013, Nylund-Gibson and Masyn 2016). In conjunction, a proportional number of the starting values (for example, 10%) that yield the largest log likelihood values in the first step should be carried over into the second step of the algorithmic optimization process (Geiser 2013). It is also recommended to record the

¹⁷ The depicted mathematical model for LTA has been adapted from Kaplan and Sweetman (2005)

number and proportion of sets of random starting values that converge to a proper solution; with a high frequency of replication of the apparent global solution across random starting values functioning to strengthen confidence in the identification of the best or 'true' ML solution (Masyn 2013). In circumstances where the best ML solution is not readily discernible from the next best log-likelihood values, the advice is to rerun these competing models using seeded starting values with consistency of the latent structures in the various iterations serving to guide decision making (Hipp and Bauer 2006).

Missing data and its management: FIML and MI

Missing data on the repeated measures can be a feature of longitudinal studies yielding unbalanced panels. And, it is the case here, in this study, where data are missing for certain neighbourhoods on indicators measuring social fragmentation across modelled time-points.¹⁸ Rubin (1976) defines data as missing at random (MAR) if the missing data pattern can be explained by observed measures and their values in the data set, and not on unobserved and unaccounted for measures or the values of the missing data themselves. Given the variables in this study (both the observed set of indicators measuring social fragmentation, and the covariates), MAR is a reasonable assumption. A corollary of the MAR assumption is that the missing data mechanism is ignorable. An ignorable missing data mechanism is important as it is an assumption of a number of model-based approaches able to accommodate missing data (Dong and Peng 2013).

Missing data introduce potential bias in parameter estimation and weaken the generalisability of the results (Dong and Peng 2013). And simply ignoring cases with missing data, for example through listwise deletion and complete case analysis, leads to information loss, decreased statistical power and increased standard errors (Little 1992). Model-based approaches to missing data are able to yield less biased estimates than simpler conventional missing data processes, for example available case analysis, single-value mean imputation, or last-value carried forward, unless the data are missing completely at random (MCAR), a stronger assumption than MAR (Dong and Peng 2013).

Two model-based approaches for missing data are **multiple** (saturated multivariate normal) **imputation** (MI), and **full information maximum likelihood** estimation using the iterative expectation-maximization algorithm (FIML-EM). Within the context of mixture models, and their direct applications, recommendations have wavered as to which of these missing data approaches to apply. For example Asparouhov and Muthén (2010) endorsed the use of MI, however, Enders and Gottschall (2011) expressed concern that the use of MI in a mixture model might differentially impact the estimates of covariate effects across latent classes.

¹⁸ In this study, there are no missing data for the neighbourhood-level predictors.

To date, recommendations related to the use of MI or FIML in mixture models have not considered the consequences of the missing data approach on class enumeration. Sterba (2016) aimed to redress this issue. Via a simulation study, the comparative model fit of a categorical latent variable, within a cross-sectional context, was examined under alternative missing data approaches, including MI and FIML. Sterba (2016, p.171) observed that MI was predisposed to reducing nonnormality, a characteristic that made it difficult to identify and recover unobserved heterogeneity, particularly “under the conditions of moderate class separation *and* large missingness proportions”. Importantly, Sterba (2016, p.172) concluded that “results should generalize to the parallel context of discriminating continuous versus categorical variability in change”.

Covariate analysis: One- and Three-step Approaches

Applications of latent class models, such as LTA, aim to ascertain a typology based on unidentified population-level heterogeneity. They can also aim to assess relationships between the modelled typology and auxiliary variables (i.e. covariates - antecedents or predictors – and, distal outcomes).¹⁹ With regards to this latter aim, two distinct approaches to *covariate analysis* have been advanced. The first method is a **single-step approach** in which covariates are directly incorporated into the latent class model. The second is a **three-step approach**. In a three-step approach: the latent class model is specified and enumerated (step-1; unconditional model), units of analysis are assigned to latent classes based on their posterior class probabilities (step-2), and associations between assigned class membership and auxiliary variables examined, in general through the specification of a multinomial logistic regression, where the latent classes are configured as categories of a now observed outcome (step-3).²⁰

The substantive difference between one- and three-step approaches relates to timing, and the matter of **when** covariate analysis is undertaken: **during** (one-step) or **after** (three-step) **latent class enumeration**. Proponents of a single-step approach assert this method yields more precise and less biased estimates of covariate effects. The basis for this assertion is that as the latent classes are enumerated jointly with the covariates this full model encompasses the most information as the covariates assist to define each class. Moreover, as a single-step approach naturally maintains the uncertainty of latent class membership within the model, conserving this probabilistic information enables the regression estimates to be modelled with classification error, a factor that can guard against any biasing of the standard errors (Clark and Muthén 2009). Equally, the single-step approach has a number of reported disadvantages. For example, the introduction of covariates directly into a latent class model can adversely impact model fit. In addition, the estimated measurement model can itself “shift in nontrivial ways with regards to the class-specific

¹⁹ In this study (and dissertation), analyses are restricted to the former type of auxiliary variables - covariates - as antecedents or predictors of the latent variables.

²⁰ Associations can also be examined descriptively using cross-tabulations; two-way tables summarising the class membership probabilities per covariate category. If this descriptive process is enacted with proportional assignment, this yields Magidson and Vermunt’s (2001) ‘inactive covariates’ method.

parameter estimates, adjusted class proportions, or even the number of classes, resulting in substantive alterations in the meaning and interpretation of the carefully constructed latent classes from the unconditional model” (Nylund-Gibson and Masyn 2016, p.783).

Shifts in the nature of the model on account of the covariates may cause concern regarding the veracity of the latent classes, as well as promote inferential uncertainty. To provide recommendations as to when to include covariates in the modelling process, Nylund-Gibson and Masyn (2016) conducted an expansive set of Monte Carlo simulations examining the influence of covariates on class enumeration. Simulations were based on latent class models with dichotomous indicators, with a range of scenarios modelled where: class frequencies (number of classes), class proportions, and sample sizes were varied. The unequivocal finding of this study was that class enumeration is most reliably undertaken in an unconditional model, and least reliable when specified through a conditional model which encompasses only an indirect path from the covariate to the observed set of indicators through the latent class (i.e. a normative one-step approach). This finding was robust, with one caveat; the enumeration of an unconditional model where the assumption of conditional independence was violated on account of omitting direct effects from the covariate to the latent class indicators, a situation that conferred an elevated risk of the over-extraction of classes, and was likely to induce bias (Nylund-Gibson and Masyn 2016).

With regards to covariate analysis, Nylund-Gibson and Masyn’s (2016) study highlighted not only when covariates should be introduced into the model process, but also **how**. In a latent class model covariates may influence not only the latent class variable, but also the observed latent class indicators. Such influences may be transmitted through an **indirect** pathway, via the latent class variable. They may also be more **direct**, conveyed through a pathway from the covariates to the indicators that bypasses the latent class variable. However, an assumption of latent class models is that the observed indicators are conditionally, or locally, independent. Omitting to specify any direct relationship between the covariates and the indicators, when there is such relationship in the population, would violate this conditional independence assumption. Moreover, as Nylund-Gibson and Masyn (2016) noted, ignoring any direct covariate effects on the indicators would impact the reliability of class enumeration in a three-step approach.

Nylund-Gibson and Masyn’s (2016, p.795) conclusion that “Latent class enumeration can and should be done without covariates” is supported by a second simulation study undertaken by Kim et al (2016), with these researchers also finding that class enumeration was robust to the exclusion of covariates from the model. Kim et al (2016) also substantiated Nylund-Gibson and Masyn’s (2016) recommendation to systematically examine direct (class-varying or class-invariant)²¹ effects on the class indicators once the

²¹ Covariates may directly affect all or some of the observed latent class indicators. Nylund-Gibson and Masyn (2016) refer to the former case as a class-invariant direct effect, and the latter case as a class-varying direct effect. Class-varying direct effects may also be referred to as ‘differential item functioning’ (DIF); see for example, Nylund (2007, pp92-93). In a neighbourhood-

covariates were modelled as predictors of class membership. In addition, though the simulation studies by Kim et al (2016) and Nylund-Gibson and Masyn (2016) were cross-sectional in nature, their principal findings regarding when and how to undertake class enumeration in the context of covariate regression, was considered to generalise to longitudinal applications.

Assignment of subjects to latent classes

An important component of the three-step 'classify-analyse' approach is the **assignment** of units (subjects; neighbourhoods) to a level of the latent class variable. Given a model, assignment is based on the estimated posterior probabilities, derived for each unit using Bayes theorem, and conditioned on the unit's pattern of responses (Bray et al 2015).

In general, there are three forms of assignment: *modal*, *pseudo-random draw*, and *proportional*. In **modal**, or maximum-probability assignment, units are assigned to the class for which they have the highest estimated posterior probability of membership (Goodman 2007). Once a unit is classified to the latent class with the highest posterior probability, its probability of being in that assigned class is one, and zero for all other classes. As such, modal assignment 'hard partitions' or crisply assigns units to classes.

Assignment with multiple **pseudo-class random draws** is also a form of **hard partitioning** (Vermunt 2010), though it aims to take account of potential membership in the array of estimated classes (Bolck et al 2004). In this approach, assignment is made by randomly sampling from the multinomial distribution of posterior probabilities. "By having multiple random samples, [units] are given a chance to flip into neighbouring classes, which gives a sense of the variation associated with the distribution" (Clark and Muthén 2009, p.12). However, unit assignment to a latent class still remains as an all or none proposition. That is assigned probabilities will be 'one' for the randomly drawn class, and 'zero' for all other classes, with regression analyses enacted once for each of the randomly drawn assignments. Modelled information is then combined across these draws in a process borrowing from the multiple imputation approach for missing data (Bray et al 2015).

The third assignment option is **proportional assignment** (Vermunt 2010). Proportional assignment applies as weights, in the regression analysis, the unit-specific array of posterior probabilities. On this basis proportional assignment is often referred to as probability weighting. Moreover, as this approach allows for units having fractional, or partial membership, in all classes it is viewed as a **soft** partitioning method.

Irrespective of which of the three assignment methods is applied, regression analyses that are based on normative three-step 'classify-analyse' approaches suffer from the attenuation of regression estimates

centred application, DIF implies that two neighbourhoods in the same social fragmentation class would have differential item endorsement probabilities. Nylund (2007) notes that depending on the 'consistency' of the DIF signal across classes, it may not "be important to incorporate it in the longitudinal model" however, "It is worthwhile to explore the effect of ignoring DIF in these models", therefore it is important that "the possibility of DIF [is] explored at each time point for [given] items."

(Bolck et al 2004, Vermunt 2010). Downward-biased estimates of the association between class membership and auxiliary variables are an artefact of the classification error introduced in a three-step approach - unless classification is perfect - when units are assigned to latent classes, with class membership effectively treated as known, or as an observed variable (Bolck et al 2004, Vermunt 2010).

Bias-adjusted three-step approaches

A number of methods have aimed to correct this attenuation by carrying over into the conditional model the model-estimated uncertainties in latent class assignment classification, weighting membership for example by the classification error (for examples see Bolck et al 2004, Clark and Muthén 2009, Vermunt 2010). Despite this attention to classification error that corrects the parameter estimates, derived standard errors can remain biased on account of the residual uncertainty regarding latent class membership and unaccounted for variance in the estimates (Bakk et al 2014). Depending on the method of assignment, modal or proportional, standard errors may be under- or over-estimated; the former functions to tighten confidence intervals and enhance the prospect of a Type-I error, while the latter may lessen statistical power (Bakk et al 2014).

A newer **bias-adjusted three-step approach**, based on likelihood theory, attends to bias of both the point estimates and standard errors, strengthening statistical inference (Bakk et al 2014). However, to date this bias-adjusted approach has only been configured for cross-sectional latent class models. While it could be applied to longitudinal models, as Di Mari et al (2016) noted, any correction of the standard errors in longitudinal applications for bias would have to account for the temporal correlation of observations and their influence on assignment; a methodological advancement still wanting.

Covariate analysis - mathematics

With regards to **LTA-covariate analysis**, to include a single, time-invariant covariate- X , and to predict its influence on latent class membership and transitions over time, the LTA model presented in equation-7 is extended as follows (equation-C3):²²

$$P(Y_i = y | X_i = x) = \sum_{s_1=1}^{n_s} \dots \sum_{s_t=1}^{n_s} \delta_{s_1}(x) \tau_{s_2}(x) \dots \tau_{s_t}(x) \prod_{m=1}^M \prod_{k=1}^{r_m} \prod_{t=1}^T p_{mk}^{I(y_m=k)} \quad [C3]$$

Where, the ρ , δ , and τ parameters are interpreted the same as in Equation-7, and 'a latent transition model with n_s dynamic latent classes is estimated based on M -categorical items measured at each of T -times for a total of MT -items, in addition to a covariate X . $Y_i = (Y_{i11}, Y_{i12}, \dots, Y_{i1M}, Y_{i21}, Y_{i22}, \dots, Y_{i2M}, Y_{iT1}, Y_{iT2}, \dots, Y_{iTM})$ represents the vector of the subject's i 's responses for all times $t=1, \dots, T$ and items $m=1, \dots, M$, where a subject response Y_{im} may take on the values $1, 2, \dots, r_m$. Also, $s_{1i}=1, 2, \dots, n_s$ is subject's i 's dynamic latent

²² The depicted mathematical models for LTA are from: Lanza et al (2010). **check for creative commons; consider further adaptation of the associated text panel*

class membership at *Time-1*, while $s_{2i}=1, 2, \dots, n_s$ is subject i 's latent status membership at Time 2, and so on. Now let $I(y = k)$ be the indicator function which equals 1 if response y equals k and 0 otherwise. In addition, X_i represents the value of the covariate X for subject i and that the value of X is allowed to relate to the probability of membership in each dynamic latent class, δ , and each transition probability, τ (Lanza et al 2010, p.*). Moreover, $\delta_{s_1}(x)$ represents a baseline-category multinomial logistic model that predicts subject i 's membership in a dynamic latent class s_1 at *time-1*, where the estimated log-odds ratio compares the odds of transitioning from the dynamic class ' n ' at time $t-1$ to the dynamic class ' n ' at *time t* relative to remaining in the dynamic class ' n_a ' at time t given the level of the covariate.

References: Appendix A

- Asparouhov, T., Muthén, B., 2010. Multiple imputation with Mplus, Mplus Web Notes, p. Retrieved from <http://statmodel.com>.
- Bakk, Z., Oberski, D.L., Vermunt, J.K., 2014. Relating latent class assignments to external variables: Standard errors for correct inference. *Political Analysis* 22, 520-540.
- Bolck, A., Croon, M., Hagenaars, J., 2004. Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis* 12, 3-27.
- Bray, B.C., Lanza, S.T., Tan, X., 2015. Eliminating bias in classify-analyze approaches for latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal* 22, 1-11.
- Clark, S.L., Muthén, B., 2009. Relating latent class analysis results to variables not included in the analysis. *StatModel, Mplus, University of California, Los Angeles*, p. 55pp.
- Di Mari, R., Oberski, D.L., Vermunt, J.K., 2016. Bias-adjusted three-step latent Markov modeling with covariates. *Structural Equation Modeling: A Multidisciplinary Journal* 23, 649-660.
- Do, C.B., Batzoglou, S., 2008. What is the expectation maximization algorithm? *Nature Biotechnology* 26, 897-899.
- Dong, Y., Peng, C.-Y.J., 2013. Principled missing data methods for researchers. *SpringerPlus* 2, 17pp, doi: 10.1186/2193-1801-1182-1222.
- Enders, C.K., Gottschall, A.C., 2011. Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal* 18, 35-54.
- Geiser, C., 2013. *Data analysis with Mplus (English ed.)*. The Guilford Press, New York, NY.
- Goodman, L.A., 2007. On the assignment of individuals to latent classes. *Sociological Methodology* 37, 1-22.
- Hipp, J.R., Bauer, D.J., 2006. Local solutions in the estimation of growth mixture models. *Psychological Methods* 11, 36-53.
- Kaplan, D., Sweetman, H., 2005. Two Perspectives on the Development of Mathematical Competencies in Young Children: An Application of Continuous and Categorical Latent Variable Modeling, in: Lissitz, R.W. (Ed.), *Proceedings of the 'Longitudinal Value Added Models of Student Performance' Conference*. JAM Press, University of Maryland, MD.
- Kim, M., Vermunt, J., Bakk, Z., Jaki, T., Van Horn, M.L., 2016. Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal* 23, 601-614.
- Lanza, S.T., Patrick, M.E., Maggs, J.L., 2010. Latent transition analysis: Benefits of a latent variable approach to modeling transitions in substance use. *Journal of Drug Issues* 40, 93-120.
- Little, R.J.A., 1992. Regression with missing x's: A review. *Journal of the American Statistical Association* 87, 1227-1237.

Magidson, J., Vermunt, J.K., 2001. Latent class factor and cluster models, bi-plots and tri-plots and related graphical displays. *Sociological Methodology* 31.

Masyn, K.E., 2013. Latent Class Analysis and Finite Mixture Modeling, in: Little, T.D. (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology*. Oxford University Press, New York, pp. 551-611.

Nylund-Gibson, K., Masyn, K.E., 2016. Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal* 23, 782-797.

Nylund, K.L., 2007. *Latent Transition Analysis: Modeling Extensions and an Application to Peer Victimization*. University of California, Los Angeles, Ca., p. 190.

Rubin, D.B., 1976. Inference and missing data. *Biometrika* 63, 581-592.

Sterba, S.K., 2016. Cautions on the use of multiple imputation when selecting between latent categorical versus continuous models for psychological constructs. *Journal of Clinical Child and Adolescent Psychology* 45, 167-175.

Vermunt, J.K., 2010. Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis* 18, 450-469.

Appendix B Syntax (Mplus) for LCA and LTA Models

For more information please consult the Mplus User Guide
**** Muthén, L.K., Muthén, B.O., 1998-2017. Mplus user's guide. Eighth Edition. Muthén & Muthén, Los Angeles, Ca.**

TITLE:
 Series of k-class LCA Models: 2001, 2006, 2011

DATA:
 File is SocialFrag_AllYrs.dat;

VARIABLE:
 Names are SSC11_ID s01_11ID s06_11ID
 mChd_01 mSPHH_01 mMarr_01 mNFHH_01 mSUR1_01 mSUR5_01
 mHOwn_01 mRImm_01 mNESp_01
 mChd_06 mSPHH_06 mMarr_06 mNFHH_06 mSUR1_06 mSUR5_06
 mHOwn_06 mRImm_06 mNESp_06
 mChd_11 mSPHH_11 mMarr_11 mNFHH_11 mSUR1_11 mSUR5_11
 mHOwn_11 mRImm_11 mNESp_11;
 Auxiliary = SSC11_ID;

Missing = all (-99);
 Usevariables =
 mChd_11 mMarr_11 mHOwn_11 mSUR1_11 mSUR5_11
 mRImm_11 mNESp_11 mSPHH_11 mNFHH_11;
 Categorical =
 mChd_11 mMarr_11 mHOwn_11 mSUR1_11 mSUR5_11
 mRImm_11 mNESp_11 mSPHH_11 mNFHH_11!
 Classes = c(4);

ANALYSIS:
 Type = mixture;

 Starts = 1000 100;

 stiterations = 100;

 lrtbootstrap = 500;
 lrtstarts = 80 40 100 50;

processors = 4;

SAVEDATA:
 File = SocFrag2011_k4.dat;

Save = cprobabilities;

OUTPUT:
 Tech1 Tech10 Tech11 Tech14 SValues;

! Text behind exclamation marks signal comments
 ! Title of series of LCA Models

! Input (data) file

! Variables in data file; here time-varying
 Suburb_IDs and the x9 x3-time varying sets of
 social fragmentation indicators

! Names and includes an ID_variable in any saved
 extension files; Useful for merging; post-
 processing
 ! Identifies any missing data
 ! Defines the observed indicators to be included
 in the model; here the x9 2011 indicators

! Defines which if any indicators are
 categorical; in this case all of the x9 social
 fragmentation indicators
 ! Names and creates the categorical latent
 variable to be modelled "c"; and, expresses how
 many classes of the latent variable are to be
 estimated - here k=4

! requests a mixture distribution (i.e. calls on
 model commands for estimating categorical latent
 variables)
 ! specifies the number of random sets of starting
 values to generate in the initial "E" step of
 estimation (here 1000), and the number of
 optimizations to use in second "M" step i.e use
 100 of the starting value sets that show the
 largest log-likelihood values in the first step
 for the second "M" step, until the convergence
 criterion is reached (default is 10 and 2)
 (conv.criterion is 0.000001 - can be changed)
 ! specifies the maximum number of iterations for
 the initial stage of the optimisation...default
 is 10
 ! requests the number of bootstrap draws (used in
 conjunction with Tech14; default is 2-100)
 ! used with the Tech14 (LRT k-1 v k test) option;
 here it is specified that for the k-1 class model
 80 random sets of starting values are used in the
 initial stage and 40 optimizations are carried
 out in the final stage; and, for the k class
 model 100 random sets of starting values are used
 in the initial stage and 50 optimizations are
 carried out in the final stage.
 ! specifies that number of processors to be used
 in the analysis for parallel computations

! Commands related to saving output and
 additional analytics
 ! Names any data file to be saved into a new data
 file; here a file that links each suburb (file
 includes the ID variable from above and a
 variable with its assigned latent class based on
 the modal posterior probabilities of membership.
 Posterior probabilities are additionally saved as
 variables for membership into all estimated
 classes
 ! asks for the posterior probabilities to be
 saved as per above

! Additional output requested:
 ! Tech1 = requests the arrays containing
 parameter specifications and starting values for
 all free parameters
 ! Tech10 = requests univariate, bivariate, and
 response pattern, as well as model fit
 information for the latent variable
 ! Tech11 = provides output related to the Vuong-
 Lo-Mendell-Rubin test (VLMRt) (k v k-1)

TITLE:

LTA Model: 2001-06-11; Measurement Invariance; k=4

DATA:

File is SocialFrag_AllYrs.dat;

VARIABLE:

Names are SSC11_ID s01_11ID s06_11ID
mChd_01 mSPHH_01 mMarr_01 mNFHH_01 mSUR1_01 mSUR5_01
mHOwn_01 mRImm_01 mNESp_01
mChd_06 mSPHH_06 mMarr_06 mNFHH_06 mSUR1_06 mSUR5_06
mHOwn_06 mRImm_06 mNESp_06
mChd_11 mSPHH_11 mMarr_11 mNFHH_11 mSUR1_11 mSUR5_11
mHOwn_11 mRImm_11 mNESp_11;
Auxiliary = SSC11_ID;

Missing = all (-99);

Usevariables =

mChd_11 mMarr_11 mHOwn_11 mSUR1_11 mSUR5_11
mRImm_11 mNESp_11 mSPHH_11 mNFHH_11;

Categorical =

mChd_11 mMarr_11 mHOwn_11 mSUR1_11 mSUR5_11
mRImm_11 mNESp_11 mSPHH_11 mNFHH_11!

Classes = c1(4) c2(4) c3(4);

ANALYSIS:

Type = mixture;

Starts = 1000 200;

stiterations = 300;

Processors = 4;

MODEL:

%Overall%

c3 on c2;

c2 on c1;

Model c1:

%c1#1%

[mChd_01\$1 - mNFHH_01\$1];

%c1#2%

[mChd_01\$1 - mNFHH_01\$1];

%c1#3%

[mChd_01\$1 - mNFHH_01\$1];

%c1#4%

[mChd_01\$1 - mNFHH_01\$1];

Model c2:

%c2#1%

[mChd_06\$1 - mNFHH_06\$1];

%c2#2%

[mChd_06\$1 - mNFHH_06\$1];

%c2#3%

[mChd_06\$1 - mNFHH_06\$1];

%c2#4%

[mChd_06\$1 - mNFHH_06\$1];

Model c3:

%c3#1%

[mChd_11\$1 - mNFHH_11\$1];

%c3#2%

[mChd_11\$1 - mNFHH_11\$1];

%c3#3%

[mChd_11\$1 - mNFHH_11\$1];

%c3#4%

[mChd_11\$1 - mNFHH_11\$1];

SAVEDATA:

file = SocFragLTA010611_9Vr_4cFreeEst.dat;

save = cprobabilities;

OUTPUT:

Tech1 Tech10 SValues;

! Tech14; = provides output related to the BLRT(k
v k-1)
! SValues = Saves parameter estimates from the
analysis that can subsequently be used as
starting values for analysis by applying them
under specified MODEL commands.

! This is an LTA Model that is to freely
estimated across the x3 time points

! Multiple Latent Variables "c"; one for each
time point, each with x4 classes

! Model command for user specific commands
! Describes the structural part of the mixture
model - here a 1st order autoregressive LTA model

! Measurement model commands; here there is one
for each latent variable - "c" (one per year) ...
there are x4 classes per year
! %c1#1% references class-1 for c1 (2001)
! [mChd_01\$1 - mNFHH_01\$1] references the mean
thresholds (in logit form) to be estimated for
each of the indicators; \$1 indicates there is one
threshold given that the indicators are binary

! As the Model is freely estimated (no imposed
measurement invariance across time), there are no
equality constraints on the mean thresholds
across "c"s (time)

! Starting values have not been used here in this
example, but can be applied (carried over) from
the LCA model ... refer to the next below
immediately below ...

```

TITLE:
    LTA Model: 2001-06-11; Structural Invariance; k=4

DATA:
    File is SocialFrag_AllYrs.dat;

VARIABLE:
    Names are SSC11_ID s01_11ID s06_11ID
    mChd_01 mSPHH_01 mMarr_01 mNFHH_01 mSUR1_01 mSUR5_01
    mHown_01 mRImm_01 mNESp_01
    mChd_06 mSPHH_06 mMarr_06 mNFHH_06 mSUR1_06 mSUR5_06
    mHown_06 mRImm_06 mNESp_06
    mChd_11 mSPHH_11 mMarr_11 mNFHH_11 mSUR1_11 mSUR5_11
    mHown_11 mRImm_11 mNESp_11;
    Auxiliary = SSC11_ID;
    Missing = all (-99);
    Usevariables =
        mChd_11 mMarr_11 mHown_11 mSUR1_11 mSUR5_11
        mRImm_11 mNESp_11 mSPHH_11 mNFHH_11;
    Categorical =
        mChd_11 mMarr_11 mHown_11 mSUR1_11 mSUR5_11
        mRImm_11 mNESp_11 mSPHH_11 mNFHH_11!
    Classes = c1(4) c2(4) c3(4);

ANALYSIS:
    Type = mixture;
    Starts = 1000 200;
    stiterations = 300;
    Processors = 4;

MODEL:
    %Overall%
    c3 on c2;
    c2 on c1;

    MODEL C1:
    %C1#1%
    [ mchd_01$1*1.26566 ] (1);
    [ mmarr_01$1*2.82762 ] (2);
    [ mhown_01$1*4.09146 ] (3);
    [ msur1_01$1*1.21030 ] (4);
    [ msur5_01$1*0.72196 ] (5);
    [ mrimm_01$1*-1.41863 ] (6);
    [ mnesp_01$1*-1.51138 ] (7);
    [ msphh_01$1*-3.03556 ] (8);
    [ mnfhfh_01$1*-2.73794 ] (9);

    %C1#2%
    [ mchd_01$1*0.11368 ] (10);
    [ mmarr_01$1*2.88866 ] (11);
    [ mhown_01$1*1.58637 ] (12);
    [ msur1_01$1*1.63233 ] (13);
    [ msur5_01$1*1.26535 ] (14);
    [ mrimm_01$1*2.00003 ] (15);
    [ mnesp_01$1*1.58592 ] (16);
    [ msphh_01$1*0.67804 ] (17);
    [ mnfhfh_01$1*0.99816 ] (18);

    %C1#3%
    [ mchd_01$1*-0.43748 ] (19);
    [ mmarr_01$1*-2.20497 ] (20);
    [ mhown_01$1*-1.89500 ] (21);
    [ msur1_01$1*-1.97964 ] (22);
    [ msur5_01$1*-1.32275 ] (23);
    [ mrimm_01$1*-0.03493 ] (24);
    [ mnesp_01$1*-1.89877 ] (25);
    [ msphh_01$1*0.37421 ] (26);
    [ mnfhfh_01$1*0.03499 ] (27);

    %C1#4%
    [ mchd_01$1*-1.20148 ] (28);
    [ mmarr_01$1*-15 ] (29);
    [ mhown_01$1*-15 ] (30);
    [ msur1_01$1*-1.04451 ] (31);
    [ msur5_01$1*-0.64211 ] (32);
    [ mrimm_01$1*0.60987 ] (33);
    [ mnesp_01$1*3.42695 ] (34);
    [ msphh_01$1*2.02488 ] (35);
    [ mnfhfh_01$1*2.06716 ] (36);

    MODEL C2:
    %C2#1%
    [ mchd_06$1*1.26566 ] (1);
    [ mmarr_06$1*2.82762 ] (2);
    [ mhown_06$1*4.09146 ] (3);
    [ msur1_06$1*1.21030 ] (4);
    [ msur5_06$1*0.72196 ] (5);
    [ mrimm_06$1*-1.41863 ] (6);
    [ mnesp_06$1*-1.51138 ] (7);

```

! This is an LTA Model however as compared to the model above, this AR1 LTA model imposes measurement invariance (equality constraints) across time. In addition, it uses the "c(1)" (2001) starting values from the freely estimated LTA model, constraining the c(2) and c(3) parameters to these estimates

! Structural model

! Measurement models ...
! Unique Starting Values - mean thresholds - one per indicator are referenced; starting values are in logit form

! Equality constraints are referenced by numeric values given in parentheses e.g. (1)

! There are x36 equality constraints that are repeated across the c(1), c(2), and c(3) sub-model commands

```

[ msphh_06$1*-3.03556 ] (8);
[ mnfhh_06$1*-2.73794 ] (9);

%C2#2%
[ mchd_06$1*0.11368 ] (10);
[ mmarr_06$1*2.88866 ] (11);
[ mhown_06$1*1.58637 ] (12);
[ msur1_06$1*1.63233 ] (13);
[ msur5_06$1*1.26535 ] (14);
[ mrimm_06$1*2.00003 ] (15);
[ mnesp_06$1*1.58592 ] (16);
[ msphh_06$1*0.67804 ] (17);
[ mnfhh_06$1*0.99816 ] (18);

%C2#3%
[ mchd_06$1*-0.43748 ] (19);
[ mmarr_06$1*-2.20497 ] (20);
[ mhown_06$1*-1.89500 ] (21);
[ msur1_06$1*-1.97964 ] (22);
[ msur5_06$1*-1.32275 ] (23);
[ mrimm_06$1*-0.03493 ] (24);
[ mnesp_06$1*-1.89877 ] (25);
[ msphh_06$1*0.37421 ] (26);
[ mnfhh_06$1*0.03499 ] (27);

%C2#4%
[ mchd_06$1*-1.20148 ] (28);
[ mmarr_06$1*-15 ] (29);
[ mhown_06$1*-15 ] (30);
[ msur1_06$1*-1.04451 ] (31);
[ msur5_06$1*-0.64211 ] (32);
[ mrimm_06$1*0.60987 ] (33);
[ mnesp_06$1*3.42695 ] (34);
[ msphh_06$1*2.02488 ] (35);
[ mnfhh_06$1*2.06716 ] (36);

MODEL C3:
%C3#1%
[ mchd_11$1*1.26566 ] (1);
[ mmarr_11$1*2.82762 ] (2);
[ mhown_11$1*4.09146 ] (3);
[ msur1_11$1*1.21030 ] (4);
[ msur5_11$1*0.72196 ] (5);
[ mrimm_11$1*-1.41863 ] (6);
[ mnesp_11$1*-1.51138 ] (7);
[ msphh_11$1*-3.03556 ] (8);
[ mnfhh_11$1*-2.73794 ] (9);

%C3#2%
[ mchd_11$1*0.11368 ] (10);
[ mmarr_11$1*2.88866 ] (11);
[ mhown_11$1*1.58637 ] (12);
[ msur1_11$1*1.63233 ] (13);
[ msur5_11$1*1.26535 ] (14);
[ mrimm_11$1*2.00003 ] (15);
[ mnesp_11$1*1.58592 ] (16);
[ msphh_11$1*0.67804 ] (17);
[ mnfhh_11$1*0.99816 ] (18);

%C3#3%
[ mchd_11$1*-0.43748 ] (19);
[ mmarr_11$1*-2.20497 ] (20);
[ mhown_11$1*-1.89500 ] (21);
[ msur1_11$1*-1.97964 ] (22);
[ msur5_11$1*-1.32275 ] (23);
[ mrimm_11$1*-0.03493 ] (24);
[ mnesp_11$1*-1.89877 ] (25);
[ msphh_11$1*0.37421 ] (26);
[ mnfhh_11$1*0.03499 ] (27);

%C3#4%
[ mchd_11$1*-1.20148 ] (28);
[ mmarr_11$1*-15 ] (29);
[ mhown_11$1*-15 ] (30);
[ msur1_11$1*-1.04451 ] (31);
[ msur5_11$1*-0.64211 ] (32);
[ mrimm_11$1*0.60987 ] (33);
[ mnesp_11$1*3.42695 ] (34);
[ msphh_11$1*2.02488 ] (35);
[ mnfhh_11$1*2.06716 ] (36);

SAVEDATA:
file = SocFragLTA010611_StrucInvT1SV.dat;
save = cprobabilities;

OUTPUT:
Tech1 Tech10 SValues;

*****

```

```
*****
```


TITLE:

LTA Model: 2001-06-11; Structural Invariance and Stationarity of Transition Probabilities; k=4

DATA:

File is SocialFrag_AllYrs.dat;

VARIABLE:

Names are SSC11_ID s01_11ID s06_11ID
 mChd_01 mSPHH_01 mMarr_01 mNFHH_01 mSUR1_01 mSUR5_01
 mHown_01 mRImm_01 mNESp_01
 mChd_06 mSPHH_06 mMarr_06 mNFHH_06 mSUR1_06 mSUR5_06
 mHown_06 mRImm_06 mNESp_06
 mChd_11 mSPHH_11 mMarr_11 mNFHH_11 mSUR1_11 mSUR5_11
 mHown_11 mRImm_11 mNESp_11;
 Auxiliary = SSC11_ID;
 Missing = all (-99);
 Usevariables =
 mChd_11 mMarr_11 mHown_11 mSUR1_11 mSUR5_11
 mRImm_11 mNESp_11 mSPHH_11 mNFHH_11;
 Categorical =
 mChd_11 mMarr_11 mHown_11 mSUR1_11 mSUR5_11
 mRImm_11 mNESp_11 mSPHH_11 mNFHH_11!
 Classes = c1(4) c2(4) c3(4);

ANALYSIS:

Type = mixture;
 Starts = 1000 200;
 stiterations = 300;
 Processors = 4;

MODEL:

%Overall%
 [c2#1](101);
 [c2#2](102);
 [c2#3](103);
 [c3#1](101);
 [c3#2](102);
 [c3#3](103);

c2#1 on c1#1(111);
 c2#1 on c1#2(112);
 c2#1 on c1#3(113);
 c2#2 on c1#1(114);
 c2#2 on c1#2(115);
 c2#2 on c1#3(116);

c3#1 on c2#1(111);
 c3#1 on c2#2(112);
 c3#1 on c2#3(113);
 c3#2 on c2#1(114);
 c3#2 on c2#2(115);
 c3#2 on c2#3(116);

MODEL C1:

%C1#1%
 [mchd_01\$1*1.26566] (1);
 [mmarr_01\$1*2.82762] (2);
 [mhown_01\$1*4.09146] (3);
 [msur1_01\$1*1.21030] (4);
 [msur5_01\$1*0.72196] (5);
 [mrimm_01\$1*-1.41863] (6);
 [mnesp_01\$1*-1.51138] (7);
 [msphh_01\$1*-3.03556] (8);
 [mnfhf_01\$1*-2.73794] (9);

%C1#2%
 [mchd_01\$1*0.11368] (10);
 [mmarr_01\$1*2.88866] (11);
 [mhown_01\$1*1.58637] (12);
 [msur1_01\$1*1.63233] (13);
 [msur5_01\$1*1.26535] (14);
 [mrimm_01\$1*2.00003] (15);
 [mnesp_01\$1*1.58592] (16);
 [msphh_01\$1*0.67804] (17);
 [mnfhf_01\$1*0.99816] (18);

%C1#3%
 [mchd_01\$1*-0.43748] (19);
 [mmarr_01\$1*-2.20497] (20);
 [mhown_01\$1*-1.89500] (21);
 [msur1_01\$1*-1.97964] (22);
 [msur5_01\$1*-1.32275] (23);
 [mrimm_01\$1*-0.03493] (24);
 [mnesp_01\$1*-1.89877] (25);
 [msphh_01\$1*0.37421] (26);
 [mnfhf_01\$1*0.03499] (27);

%C1#4%

! This is a similar model to the one above (LTA AR1 model with measurement invariance) however here Stationarity is being imposed on the propensity for transitions across time points

! Addition of equality constraints in the structural part of the model that function to constrain transition probabilities to be the same across time

[Means] ... appear in square brackets

Variances ... appear unbracketed

! Measurement model ...

! Unique Starting Values - mean thresholds - one per indicator are referenced; starting values are in logit form

! Equality constraints are referenced by numeric values given in parentheses e.g. (1)

! There are x36 equality constraints that are repeated across the c(1), c(2), and c(3) sub-model commands

```
[ mchd_01$1*-1.20148 ] (28);
[ mmarr_01$1*-15 ] (29);
[ mhown_01$1*-15 ] (30);
[ msur1_01$1*-1.04451 ] (31);
[ msur5_01$1*-0.64211 ] (32);
[ mrimm_01$1*0.60987 ] (33);
[ mnesp_01$1*3.42695 ] (34);
[ msphh_01$1*2.02488 ] (35);
[ mnfhf_01$1*2.06716 ] (36);
```

MODEL C2:

```
%C2#1%
...
%C2#2%
...
%C2#3%
...
%C2#4%
...
```

MODEL C3:

```
%C3#1%
...
%C3#2%
...
%C3#3%
...
%C3#4%
...
```

SAVEDATA:

```
file = SocFragLTA010611_StrucInvTlSV_Stationarity.dat;
save = cprobabilities;
```

OUTPUT:

```
Tech1 Tech10 SValues;
```

! Unique Starting Values imposed to equality across time have been annotated as per the previous example

Appendix C Visualisations: 2011 LCA indicators in continuous form

<<<
Scatterplot
Correlation Matrix
of the n:9 2011
Census Indicators

Appendix D LCA Profile Plots: *k4-to-k6*; 2001, 2006, 2011

Figure-X: Conditional item response probabilities for 2006 LCA Models k=4 to k=6. These conditional responses indicate the probability of a value above the median within a class. Classes are nominal (unordered) and unlabelled

Figure-X: Conditional item response probabilities for 2006 LCA Models k=4 to k=6. These conditional responses indicate the probability of a value above the median within a class. Classes are nominal (unordered) and unlabelled

Appendix E Chloropleth maps for LCA $k=4$ -to- $k=6$ solutions; 2001, 2006, 2011 and the final LTA solution

2001 LCA $k=4$ through to $k=6$... SSCs assigned on the basis of highest posterior probabilities

*** ignore numeric class order *** colour codes reference corresponding classes across $k+1$ models and across time

2006 LCA k=4 thru k=6 ... SSCs assigned on the basis of highest posterior probabilities

2011 LCA k=4 thru k=6 ... SSCs assigned on the basis of highest posterior probabilities

Final LTA solution (2001, 2006 and 2011; $k=4$ with SSCs assigned based on highest posterior probabilities)