A Cohort Location Model of household sorting in US metropolitan regions

Hossein Estiri
Harvard Medical School, USA

Andy Krause
University of Melbourne, Australia

Abstract

In this paper we propose a household sorting model for the 50 largest US metropolitan regions and evaluate the model using 2010 Census data. To approximate residential locations for house-hold cohorts, we specify a Cohort Location Model (CLM) built upon two principle assumptions about housing consumption and metropolitan development/land use patterns. According to our model, the expected distance from the household's residential location to the city centre(s) increases with the age of the householder (as a proxy for changes in housing career over life span). The CLM provides a flexible housing-based explanation for household sorting patterns in US metropolitan regions. Results from our analysis on US metropolitan regions show that house-holds headed by individuals under the age of 35 are the most common cohort in centrally located areas. We also found that households over 35 are most prevalent in peripheral locations, but their sorting was not statistically different across space.

Introduction

Household mobility is the 'engine of change' in metropolitan areas (Clark et al., 2014). Between 2012 and 2013 more than 35 million people changed their residential location in the USA (Ihrke 2014). Data from the Current Population Survey (CPS) (King et al., 2010) show that, on average, about 75% of mobility between 1999 and 2010 in the USA occurred because of housing-related or household-related reasons, and almost 60% was within counties – most likely within metropolitan areas. Classic

Corresponding author:
Hossein Estiri, Harvard Medical School, Massachusetts General Hospital, Laboratory of Computer Science, 50 Staniford Street, Suite 750, Boston, MA 02114, USA.
Email: hestiri@mgh.harvard.edu

residential mobility theories attribute these intra-metropolitan moves to dissatisfaction with current residential location, owing to the mismatch between household's actual and desired housing consumption levels (Brown and Moore, 1970; Clark and Onaka, 1983; Speare, 1974). According to this approach, households move across residential locations to establish and maintain equilibrium between their actual and desired housing consumption. The definition of housing in this context goes beyond the actual residential unit, embracing a bundle of housing services – e.g. housing type, size, structure, proximity to schools and grocery stores. Housing consumption, therefore, refers to: 'housing behavior driven by housing needs and preferences over and above the basic need for shelter' (van Ham, 2012: 50).

While prior research has established an ample understanding of residential mobility and housing choice mechanisms, we know somewhat less about the spatial implications of these choices across metropolitan areas (Clark and Maas, 2015; Hedman et al., 2011). The distribution of the bundle of housing services often shapes specific spatial land use patterns in metropolitan areas across different contexts. These patterns do not change rapidly – an example of such patterns is the suburbanisation in the USA. In contrast to the distribution of housing services, household characteristics change in relatively short periods of time and so change their housing needs and desires. Assuming no limitations to residential mobility, household changes, thus, are the main causes of residential relocation across metropolitan areas, creating population sorting effects.

Building upon a recent population sorting model, the 'phasic' model (Estiri et al., 2015), this paper uses housing consumption and residential mobility theories and Census data from the largest 50 US metropolitan regions to develop and evaluate a cohort-based population sorting model, the Cohort Location Model (CLM), for the spatial distribution of households within metropolitan areas. CLM's contributions are twofold. First, it adds to the lifecycle-versus-life course debate by showing that age still matters in predicting housing consumption behaviours. The model shows that there are age-based sorting effects happening across metropolitan regions of the USA. Second, the CLM contributes to our knowledge on the interconnections between metropolitan land use patterns and households' housing consumption that influence population sorting patterns.

Why families move

Dissatisfaction (or satisfaction) with residential locations, the key stimuli for intra-metropolitan residential mobility, happens through interactions between household characteristics and aspirations and attributes of their dwelling units and neighbourhoods. Therefore, the desire to move is initiated by changes in households' socio-demographic characteristics or in response to transformations in their residential location. Contextual characteristics such as locational amenities, accessibility and socio-economics of the neighbourhood also influence households' (dis)satisfaction with their residential location (Dieleman, 2001). For example, households often relocate to adjust to (or improve) their neighbourhoods (Durlauf, 2004). Social and place-specific inter-relationships also play an important role in creating a sense of identity for households and individuals (Winstanley et al., 2002), especially in later stages of their housing careers (Clark and Coulter, 2015). In addition, conditions of the local and regional housing markets are also important in defining demand for housing (Dieleman et al., 2000; Ioannides and Zabel, 2008; van der Vlist et al., 2002) and therefore influence mobility behaviours.

Yet, the effect of locational attributes on intra-metropolitan residential mobility behaviours are secondary, compared with household- and housing-related inducements (Cervero and Duncan, 2006; Clark and Coulter, 2015; Ihrke, 2014; Lee et al., 1994; Nechyba and Walsh, 2004; Waddell et al., 2007).

At the intra-metropolitant level, residential mobility behaviours are mainly influenced by household and housing factors (Clark and Dieleman, 1996; Clark et al., 2006). For example, household size is positively correlated with the need for residential space and plays an important role in influencing households' residential satisfaction and initiation of mobility decisions (Clark and Huang, 2003). A larger family size increases the probability of choosing a large unit (Ewing and Rong, 2008). For example, for households who live in crowded housing units, additional children may trigger dissatisfaction with their current housing unit and initiate a potential relocation (de Groot et al., 2011). Therefore, high rates of residential mobility are expected to be associated with areas characterised by small dwellings and/or around the time of a child's birth (Clark et al., 2002).

Home tenure status is also one of the most important factors in residential relocation of households. In general, homeownership can significantly reduce mobility (Clark and Huang, 2003; Lee et al., 1994). Higher mobility rates are consequently more likely to occur in markets/neighbourhoods with a large rental stock and a younger population (Dieleman et al., 2000; van der Vlist et al., 2002).

The lifecycle approach to residential mobility

Determinants of residential mobility, such as household characteristics and their housing needs and expectations, change over time and across the life of households. Therefore, householder age has been used as a proxy for predicting residential mobility behaviors (Clark, 2013). The lifecycle approach was common in residential mobility studies conducted from the 1950s to the 1970s (Geist and Mcmanus, 2008; Kulu and Milewski, 2007). According to the lifecycle approach, while other socioeconomic factors are useful markers in the mobility process, changes in the lifecycle stage are the most significant inducers of the mobility of households (Simmons 1968). These changes act as 'triggers' generating 'disequilibrium' in housing consumption and the subsequent relocation (Clark and Dieleman, 1996). The basic principle is that households move between lifecycle stages and they re-evaluate the characteristics of either their current neighbourhood or current housing unit based on new standards (Lee et al., 1994).

Clark and Onaka (1983) argue that residential movements that correspond with lifecycle changes encompass a substantial share of intra-metropolitan moves, independent of housing unit characteristics. The classification of lifecycle stages primarily relies on marriage and the presence of children. For example, Abou-Lughod and Foley (1960) described a household's lifecycle and its associated mobility behaviour in four stages, including: pre-child, childbearing and childrearing, child-launching and post-child (Abu-Lughod and Foley, 1960). Frey (1978) and Speare (1970) also classified family status in five subgroups corresponding to family lifecycle attributes such as marital status, age and the presence of children (Frey, 1978; Speare, 1970).

In general, in the lifecycle approach, age is negatively associated with the probability of moving (Boehm and Ihlanfeld, 1986; Landale and Guest, 1985; Lee et al., 1994). Recent research has verified that this trend still holds (Beige and Axhausen, 2012; Clark et al., 2014; Coulter and van Ham, 2013;

Wulff et al., 2010). This inverse relationship between the age factor and mobility has been one of the most consistent similarities among multiple residential mobility studies (Quigley and Weinberg, 1977). Simmons (1968), in a synthesis of mobility literature, concluded that of all moves within a lifecycle stage, about 20% occur for ages under 10, almost 60% happen between the ages of 10 and 40, and 20% after the age of 40 (Simmons, 1968).

Application of a lifecycle-inspired approach in modelling household sorting patterns

Several demographic and contextual factors determine mobility behaviours of households. According to the lifecycle approach age can be used as a proxy for changes in households to approximate general residential mobility patterns. For example, prior research has shown that younger households are generally expected to be more mobile, a pattern that will stabilise over time during a given household's housing career (Beige and Axhausen, 2012; Wulff et al., 2010). Research has also suggested that areas with higher proportion of rental dwellings and smaller housing units are likely to have higher mobility rates (Clark and Morrison, 2012; Dieleman et al., 2000; van der Vlist et al., 2002). In the USA, with its relatively high degree of suburbanisation, areas closer to city centres often have a greater proportion of rental dwellings and smaller housing units (compared with suburban areas), and therefore are expected to have higher mobility rates. Therefore, younger adults, who also have a higher degree of mobility, are expected to live closer to inner city locations.

Based on the lifecycle approach, Estiri et al. (2015) developed a model to project household distribution patterns across major US metropolitan regions based on three phases of the household lifecycle. They

showed that there is a clear distinction between phase one households (householder between ages of 15 and 34) and phase two and three households (ages 35 and above), where phase one households are more likely to reside closer to city centres. They constructed their model based on two hypothetical probability distribution functions that differentiate three phases for households' housing career over their lifecycle. They argued that households in phase one have lower housing consumption and higher mobility. As households move along their lifecycle, housing consumption increases and residential mobility decreases between the ages 35 and 59 (phase two). In phase three (after age 60) households are slightly more likely to move because of changes in their housing needs. Based on this classification, the authors argued and illustrated that younger US households (phase one) are more likely to live closer to city centres. In contrast, they demonstrated that phase two and phase three households often locate in suburbs where they can find higher levels of the bundle of housing services (Estiri et al., 2015).

Relaxing the phasic model

Despite its application as a proxy for changes in household characteristics, the lifecycle approach has been heavily criticised for overly simplifying households' mobility behaviours, being deterministic and yet inconsistent in categorising lifecycle stages, and not fully embracing the demographic transitions in modern households (Bailey, 2008; Clark and Withers, 2007; Geist and Mcmanus, 2008; van Ham, 2012). As a result, the lifecycle was gradually substituted with the lifecourse approach to residential mobility, which envisions household career as a sequence of parallel and interrelated transitions of life events (e.g. marriage, having children, work and education, etc.)

(Clark and Dieleman, 1996; Clark and Withers, 2007; Kulu and Milewski, 2007; van Ham, 2012).

This paper evaluates whether the age of the householder can still be a reliable predictor of changes in a household's housing career and be used to approximate the location of residence for households in the 50 largest US metropolitan areas. We conduct this evaluation by relaxing the phase-based assumptions of the phasic model. The phasic model characterises households' housing careers in three phases based on grouping households by the age of the householder. According to the phasic model, younger adults (phase one) have lower levels of housing consumption and thus are more flexible to relocate. As they move across their life span, their housing consumption increases quickly until it peaks in phase two and then begins to fall slowly. In phase two households are very reluctant to relocate. Changes in housing consumption and residential mobility in phase three happen slowly. As households enter phase three, changes in their socio-demographics (specially, household size) and their locational preferences may lead them to donwsize and relocate (Estiri et al., 2015).

In this paper, we make an adjustment to the phasic model by relaxing the assumption of identifying any threshold for phases and approach households' life span as a continuous spectrum. Even though we still adapt the three main assumptions for patterns of change in housing consumption from the phasic model, we do not make any assumptions for when such changes could happen. We introduce our model as the Cohort Location Model (CLM).

The Cohort Location Model (CLM)

In the CLM, we argue that changes in households' housing careers (i.e. housing consumption, residential mobility and location choices) across the life span happen in a continuum and therefore we hypothesise that such changes could create spatial sorting effects within metropolitan areas. Our approach approximates household housing careers over a household's life span, which we conceptualise as a continuous variable that can be broken into as many segments as empirically testable. With this approach, modelling housing careers becomes more flexible, as any specific pattern can be approximated over multiple intervals based on data availability – i.e. the life span can be broken into years, decades, stages (as in life-cycle studies), or phases (as in the phasic model). Similar to the phasic model, we proxy a household's life span with age of the householder.

Also, similar to the phasic model, in CLM we assume that owing to the predominance of low-density residential development ('suburban') in most US metropolitan regions, the provision of housing services[1] generally increases with distance from city centre(s). By using the term 'city centre(s)' we imply the application to both mono-centric and multi-centric cities – we conventionally substitute city centre(s) with CBD in this article. The increase in the provision of housing services may also be reflected in property values. Yet, there are distinctions in the type of housing services distributed between the suburbs and the metropolitan fringe that make each location favourable for specific groups of households.

The general assumption in the CLM is that because consumption of housing services increase over a household's housing career (which we approximate here with the age of the householder), the youngest cohort of households is more likely to live closest to the city centre(s) where fewer housing services are offered, and vice versa. We hypothesise that there is a broad spatial order to this pattern, as illustrated in Figure 1.
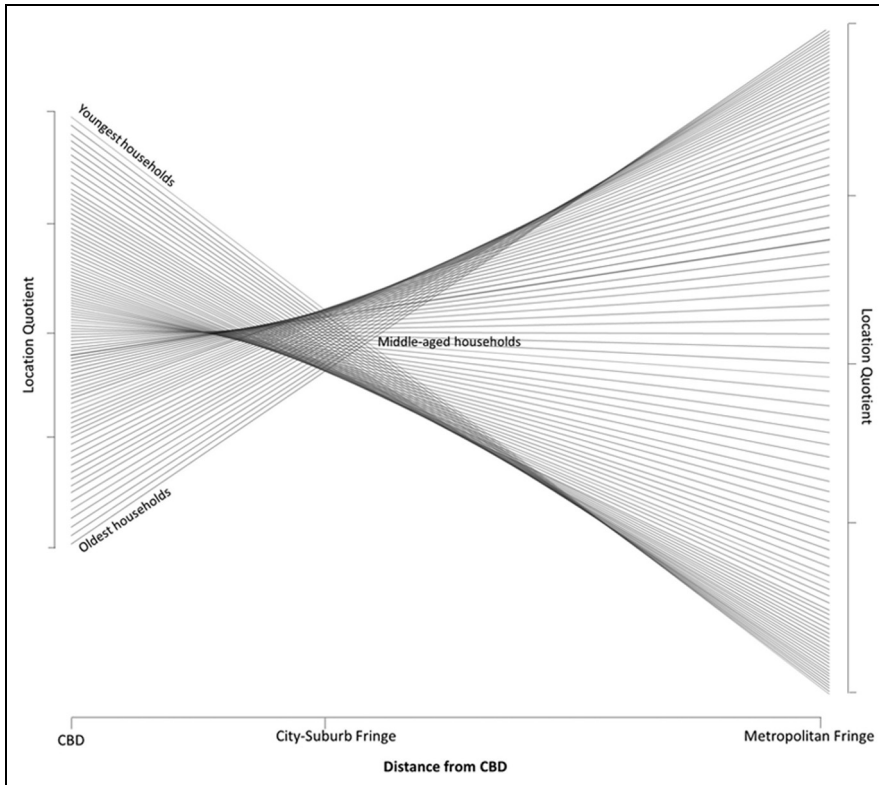
Figure 1. A linear approximation for the probability (location quotient) of households' residential location across metropolitan regions.

The X axis in Figure 1 represents a two-dimensional profile of the metropolitan region, from the city centre (conventionally called the CBD) or a significant subcentre, to a hypothetical border between the city and the low density suburban areas (conventionally called the city–suburb fringe), and to the metropolitan fringe (i.e. where the metropolitant region ends). The Y axis represents the relative location quotient for each of the various cohorts. The location quotient represents the relative likelihood of a cohort living in a certain distance from the CBD. More specifically, it measures the proportion of households of a certain cohort at a given distance to the proportion of households of that cohort in the entire region. Changes in the location quotient for each cohort across

metropolitan space are presented by a line, which is more compact on the left and more dispersed on the left – reflecting the relative compactness of central city locations compared with the metropolitan fringe. According to the model, as the age of the householder increases, the probability of the household living close to the city centre(s) decreases. The overall trend is as follows:

- At the city centre(s), the sorting pattern is more distinguishable for younger households (who are the prominent age groups at central city locations) and older households (who, based on the assumptions of this model, are the least expected age groups), than for the middle-aged householders – i.e. at the

city centre(s), it should be easier to distinguish among the ages of, for example, 24 and 25, or 78 and 79, than 42 and 43.

- If modelled linearly, the trend lines (i.e. the linear approximations of the probability functions) begin to converge somewhere near the city–suburb fringe. The CLM assumes that middle-aged households are more likely to live in proximity to where suburbs emerge outside of central city locations. The middle-aged households in this geographic domain are followed by older and then younger households.

- Between the suburbs and the metropolitan fringe, the likelihood patterns exhibit gentler transitions. At the metropolitan fringe, the residential likelihood pattern is exactly opposite to the pattern at the city centre(s): older households are the most likely to live and difficult to distinguish, middle-aged households are less likely to reside but easy to distinguish, and younger households are least likely to live and also difficult to distinguish.

We argue that the model presented here is empirically testable with different time intervals – time intervals can be defined flexibly based on data availability and granularity of the data. Using 2010 US Census data, we evaluated this model for eight aggregate trend lines illustrated in Figure 1, representing 10-year age groups or cohorts. This aggregation is due to data availability.

Method

To test the assumption discussed above we analysed household location data from the 50 largest (as of 2013) Core Based Statistical Areas (CBSAs) in the USA. A list of the 50 largest CBSAs and their 2010 population is provided in Appendix 1. Together, the selected CBSAs had a population of 166,033,092 in 2010 – comprising 54% of the total population in the USA. We derived data on household age and location (at the census block level) from the 2010 Census SF1 data set. In addition to the generalisability resulting from the population coverage, the 50 largest metropolitan regions were selected in previous studies that involved data and computations at similar spatial scales (for example, see Denton and Massey, 1991; Gober et al., 2013; Hunt and Balachandran, 2015; Lang, 2002; Lichter et al., 2015; Markusen and Schrock, 2006; Sivak, 2008).

Data preparation

We began by collecting geographic data (i.e. census block shape files) and SF1 data (flat files) from the USA. Census's FTP site[2] for each state which contains at least one county in one of the 50 largest CBSAs. From the geographic data, we isolated all geographic records at the census block level. These data include a unique identifier that links to the SF1 data, a field noting which CBSA each block is located in, if any, as well as fields indicating the latitude and the longitude for the census block centroid. Using the CBSA field within the geographic data, we removed all census blocks not located within the largest 50 CBSAs.

From the SF1 data we extracted the responses to question P22 that breaks down households into family and non-family households and by the age of the head of household. Because responses to this question are provided in 10-year cohorts (except for the 55 to 59 and 60 to 64 groups), we develop a consistent 10-year cohort categorisation for the analysis. As the family versus non-family designation does not play a role in this analysis, we combined these figures within each of the 10-year age cohorts, leaving us with eight age groups for each census block. We then merged the CBSA

identification number and latitude and longitude value from the geographic data to the SF1 P22 data, using the unique census block identification number. Census blocks with no households were eliminated at this stage. This merged data set is hereafter referred to as the 'Household Location Data' or HLD – we have made this data set publicly available online.[3]

We applied a systematic approach to identify centre and subcentre(s) for each metropolitan region. For each of the 50 CBSAs, the exact latitude and longitude points for all centre and subcentre(s) locations were obtained via Google Maps. CBSA centres are the first city to appear in the CBSA name – e.g. Phoenix in the Phoenix-Mesa-Scottsdale CBSA. Subcentres are the subsequent locations within the CBSA names – Mesa and Scottsdale, for example. Non-municipal subcentre names (such as Kenosha County or Northern New Jersey) were removed from this analysis as deriving a single centroid point is not possible. Using the HLD, the great-circle distance from each census block centroid to its corresponding CBSA centre and subcentres, if any, was then computed using the Haversine formula. The minimum distance from each census block to a centre or subcentre(s) was recorded and added to the Household Location Data.

Computation of the distance variable

As we intend to compare household location by distance over a variety of metropolitan areas – many vastly different in size – we began by standardising the distance measurements. First, we removed all census blocks with a minimum centre/subcentre(s) distance of greater than 60 miles as these likely represent highly rural areas not within the economically functional range of the metropolitan region. The 60-mile cutoff was based on a natural break in the data. Next, we standardised all distances to be a fraction

of the greatest minimum centre/subcentre(s) distance of the remaining census blocks. For instance, if the farthest census block in a given CBSA is 25 miles from its nearest centre/subcentre(s), and block X is 5 miles from its nearest centre/subcentre(s) then block X has a standardised distance of 0.2 (5/25). Doing so converts all distances to a value between 0 (at the centre/subcentre) and 1 (at the metropolitan fringe). Standardised distances were rounded to two digits. Figure 2 shows a map produced from the Atlanta-Sandy Springs-Marietta CBSA, an example of the data we compiled for each of the 50 CBSAs.

Computation of the location quotient

The Cohort Location Model described above theorises that younger households are more likely to locate near the city centre(s) while older households are more likely to locate near the fringe. As overall household counts are not evenly distributed by the age of the householder, simple counts of households at each location will not result in fair comparisons. Therefore, to evaluate the location choices of a given age of a household, we computed location quotients for each age group at each 1/100 of the standardised distance. Location quotients are particularly useful in this regard because they allow for differences in the underlying distribution of households by age to be held constant while comparison across distance – the focus of this research – are made. Location quotients are calculated as:

$$\frac{\frac{HH_{ij}}{HH_j}}{\frac{HH_i}{HH}} \quad (1)$$

where HH is the total count of households, i is the 10-year cohort of the head of household and j is the standardised distance rounded to the nearest 0.01.
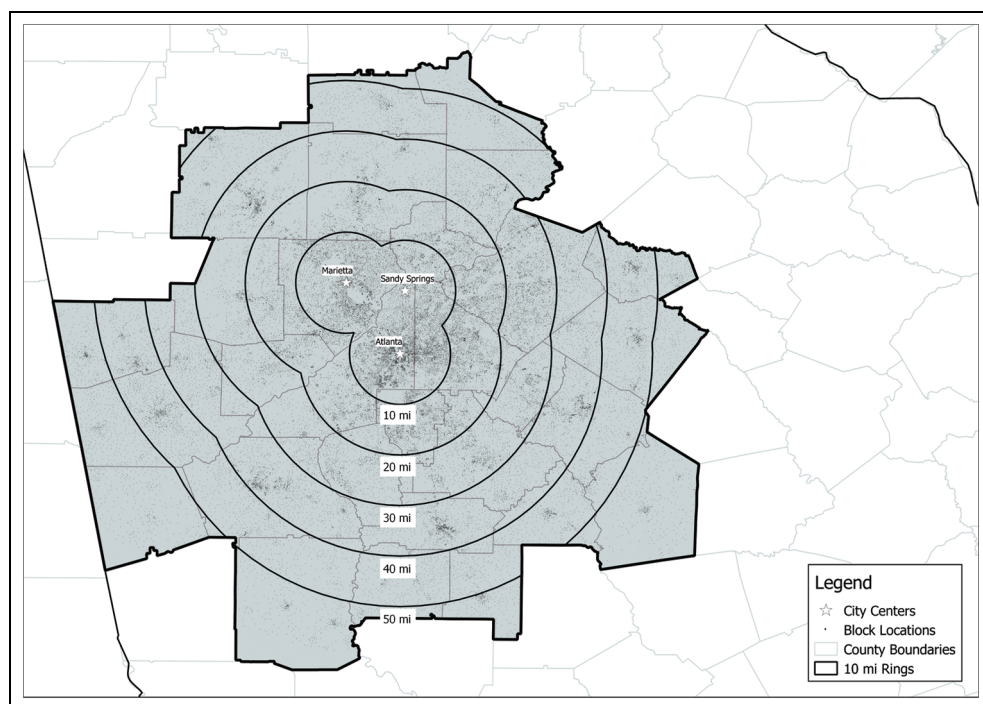
Figure 2. Centres' location and distance computation for the Atlanta-Sandy Springs-Marietta CBSA.

A location quotient of 1 indicates that a given age cohort is represented at a given distance in the same proportion as that age cohort is represented in the entire metropolitan area. Location quotients (LQs) less (greater) than 1 indicate under(over)-representation in a given area or at a given distance.

To test the applicability of the theoretical model developed above, we calculated location quotients for each age group at each of the 100 standardised distances. We then plotted these location quotients. Owing to the high variability at some distances caused by low sample sizes of households, we smoothed the overall trend lines using a locally weighted regression technique (LOWESS or LOESS, with a smoothing factor of 0.1). The LOESS-smoothed lines present a more reasonable trend regarding the changes in the frequency of household location by age over distance. Location quotient

analyses are performed on the combined 50 CBSA data set as well as individually on each CBSA. The results of these analyses are discussed below.

Results

Results of our analyses on the 50 CBSAs support the assumptions we made in the CLM. For each age group, Figure 3 illustrates two patterns of the distribution of households aggregated across the 50 US metropolitan regions.

The dashed lines represent a linear regression line along with the LOESS-smoothed pattern lines (continuous lines with a confidence interval) for each age group. The dots on each of the age group plots represent the aggregated location quotients of the households' residence by distance from the city centre(s). To improve visualisation, we
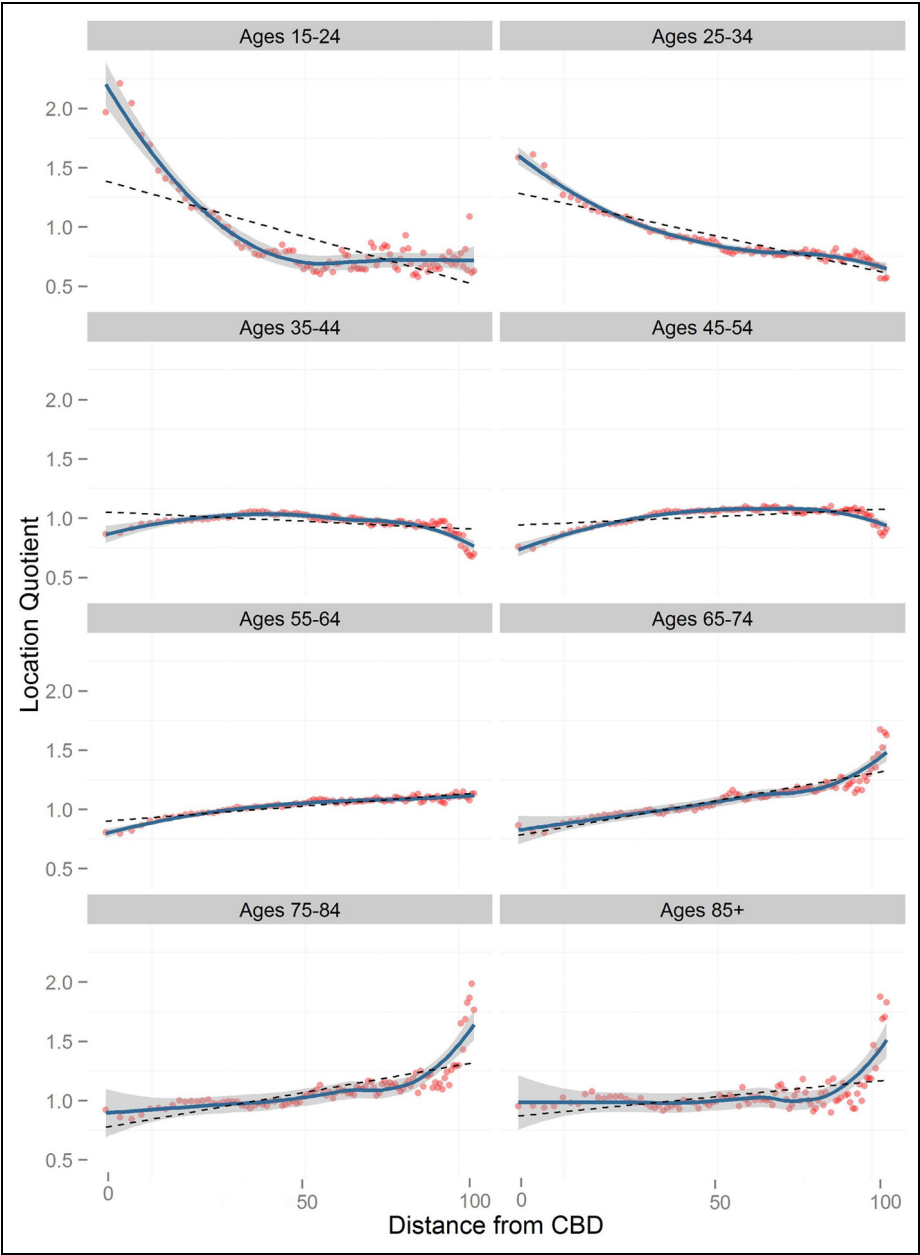
Figure 3. Plotting trend lines of the location quotients for the distribution of households with age of householder aggregated across the 50 US metropolitan regions.

applied a jitter effect to the scatterplot of the location quotients that are plotted under the trend lines.[4]

Figure 3 clearly demonstrates that, on average as we move from the youngest age group (15 to 24 years) to the oldest (>85
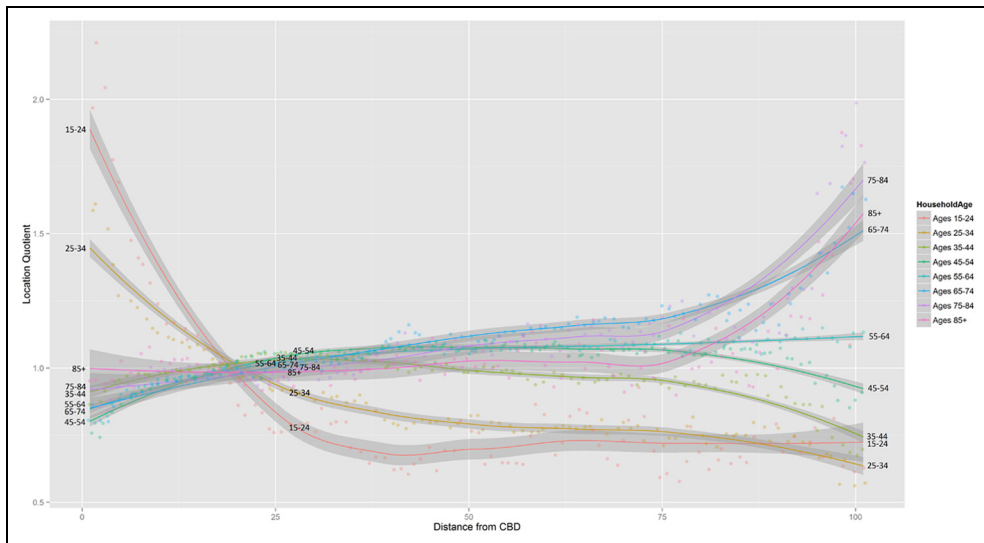
Figure 4. Overlap of LOESS-smoothed lines for the eight age groups.

years) the slopes for trend lines change from negative to positive. That is, as the age of the household increases, the probability of the household living in farther distances from the city centre(s) increases. A special distinction in slope change happens between households whose householders are younger than 34 years (with a negative slope) and older households (with a positive slope). For the youngest group, there seems to be high variability across all distances from the city centre(s) – as wider confidence intervals suggest. Between the ages of 25 and 64, variability is considerably lower. However, the variability increases again for ages older than 74, especially at the city centre(s). These instances of high variability are possible indications of a relatively smaller population for the particular age group at a given distance compared with the population of other age groups across all metropolitan regions. In addition to identifying variability in trend estimates, confidence intervals plotted in Figure 3 are useful to statistically compare the trend lines.

Differences in smoothed pattern lines

Though the linear regression trends exhibit the general patterns, the LOESS-smoothed lines provide a better illustration of spatial variability (and the overall trends) – because a smoothed regression line often fits the data better. Figure 4 combines the smoothed patterns and their respective standard errors for all age groups. In other words, Figure 4 overlays the individual plots for each age group from Figure 3 in one graph for comparative purposes. To identify the order of presence, we marked age groups at the city centre(s), the metropolitan fringe and after the patterns converge (or the beginning of the suburbs).

Looking at the order, the figure shows that at the city centre(s) the 15-to-24 cohort (i.e. with a head age 15 to 24) has the highest location quotient, by a large margin. That is, households from this cohort are significantly more likely to reside at the city centre(s), compared with their distribution in farther distances and to all other households. Households whose householders are between 25 and 34 years old (the 25-to-34

cohort) also show increased likelihood of locating in the most central areas of cities The increased location quotients for the two youngest age groups are statistically significant from the trends of other groups.

Interestingly, the oldest households (85 +) are show a higher likelihood than all other 35 + year old household to be located in the city centre. Their location quotient at the city centre does remain less than 1 meaning that even though the 85 + cohort is more likely than the households between 35 and 84 to live in the city centre(s), they are still more likely to reside farther from the city centre(s) – i.e. the location quotient increases with distance from the CBD. From a statistical standpoint the confidence intervals overlap for all 35 + cohorts, suggesting that the differences are not statistically significant for age groups above 35, as we hypothesised in our model.

As we move across the metropolitan region and into the suburbs (from a standardised distance from about 25 to 40 miles the CBD in Figure 4), the two youngest cohorts (15-to-24 and 25-to-34) immediately distinguish themselves as the cohorts with the lowest location quotients. This difference is also statistically significant, according to the confidence intervals. In contrast, middle-aged households become the majority groups in the suburbs, or from about from the CBD in Figure 4. Different from the order in city centre(s), households from the 45-to-54, cohort – who were the the least probable in the CBD – are the most likely households to be found in the suburbs, followed by the 35-to-44 and 55-to-64 cohorts. However, most of these patterns are not statistically different for the households older than the 35-to-44 cohort. We also hypothesised in our model that it could be difficult to distinguish between the two middle-aged cohorts (the 35-to-44 and 45-to-54 cohorts) in the suburbs because of their high concentrations in these locations. After the middle-aged

households, the likelihood of residence in the suburbs decreases by age of the householder – note that the youngest households (15 to 34) still have the lowest probabilities of living in the suburbs.

Moving towards the metropolitan fringe (ex-urban areas), households sort from the highest to the lowest location quotient by cohort almost exactly as we hypothesised: 75-to-84, 85 + , 65-to-74, 55-to-64, 45-to-54, 35-to-44, 15-to-24, and 25-to-34. Also, as our model suggested, it is quite possible to distinguish middle-aged households from their older and younger peers in these locations. From the suburb to the metropolitan fringe (standardised distance X to Y), the location quotient decreases for households whose householder is younger than 54, and vice versa. The most steep increasing trends across these grographies belong to the cohorts with householders over 65 years old. For these cohorts, however, the confidence intervals overlap at the metropolitan fringe, meaning that the difference in their location quotient patterns is not statistically significant compared with households in the 85 + cohort. Similarly, confidence intervals overlap for the younger cohorts (with householders younger than 44 years old). Part of these overlaps at the metropolitan fringe can be due to the big confidence intervals (large standard errors) of the younger and older households, which in turn can be due to their relative small population this far from the city centre(s).

Changes and differences in linear patterns

To evaluate our proposed CLM, we also overlapped linear regression lines from Figure 3 and added standard errors to create Figure 5. As the model introduced in Figure 1 was based on a linear approximation of the patterns, linear regression lines provide the best visuals to evaluate our model.

Although the linear regression lines do not project variability as well as smoothed
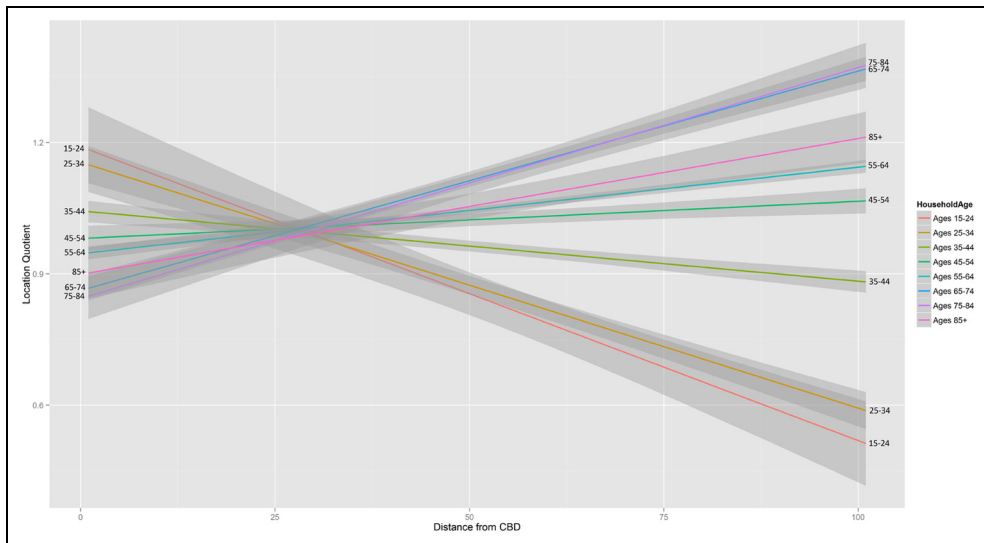
Figure 5. Overlap of linear regression lines for the eight age groups.

trend lines do, they are effective in illustrating the overall increasing/decreasing patterns. The trend lines illustrated in Figure 5 almost mirror our hypothetical model in Figure 1. According to Figure 5, households that live in the city centre(s) sort from the youngest cohort (15-to-25), with the highest location quotient, to the second oldest cohort (75-to-84) with the lowest location quotient – the only exception is that the 85 + cohort ranks before the 65-to-74 cohort. From the eight cohorts, four clusters of cohorts are statistically distinguishable. Even though the 15-to-24 cohort has the highest location quotient for residence in the city centre(s), the difference between location quotients for this group and the 25-to-34 cohort is not statistically significant throughout the metropolitan space. Similar to these young household groups are households from cohorts 65-to-74 and 75-to-84. Therefore, the eight cohorts can be combined in city centre(s) into four clusters of: (1) youngsters (the 15-to-24 and 25-to-34 cohorts), who have the highest location

quotient, (2) the 35-to-44 cohort, (3) the 45-to-54 and 55-to-64 cohorts, and (4) the older cohort (65 +) who have the lowest location quotient in the city centre(s).

At the metropolitan fringe the sorting pattern that we observed in the city centre occurs in reverse order. Households from the 75-to-84 cohort have the highest location quotient – and therefore, the probability to reside – at the metropolitan fringe. After this cohort are households from the following cohorts: 65-to-74, 85 + , 55-to-64, 45-to-54, 35-to-44, 25-to-34, and 15-to-24, in order from the second highest to the lowest location quotient at the metropolitan fringe. Except for the two youngest (15-to-24 and 25-to-34) and two of the older (65-to-74 and 75-to-84) cohorts, the regression lines differ statistically for households whose householders are between 35 and 64. As a result, the four-clusters pattern in the city centre(s) extends into a five-clusters pattern at the metropolitan fringe. These clusters quickly distinguish themselves right after the regression lines converge – which means that these

different patterns are expected to be visible almost throughout the area from the city–suburb fringe to the metropolitan fringe.

Conclusion and discussion

Recent research has shown evidence of the role of population age in shaping the urban spatial patterns (Estiri et al., 2015; Moos, 2015). Building upon a recently proposed model of population distribution patterns, the Cohort Location Model presented in this paper describes a sorting mechanism for the distribution of households across US metropolitan areas by age of the householder.

We developed this model based on two assumptions. First, we expect housing consumption to increase over a household's housing career in accordance with the increase in the household's housing needs and subsequent change in housing preferences. Using the age of the householder as a proxy for the household's housing career, we assume that as the age of the householder increases, households' housing consumption increases with a logarithmic pattern. Second, we expect that because of the prevailing suburban development pattern in US metropolitan regions higher levels of housing services are offered when the distance from the city centre increases. Intersecting the two assumptions, our model envisions a metropolitan sorting effect in which younger households are more likely to live closer to the city centre(s) and older households are more likely to live closer to the metropolitan fringe.

We evaluated our CLM using data from the 50 largest metropolitan regions in the USA. Results showed that the slope of location quotient patterns changes from negative (i.e. a decrease in the probability of residence as the distance from city centre(s) increases) to positive as the age of the householder increases. This transition in the slope of the trend lines happens somewhere around the

age of 35, which is consistent with the phasic model's (Estiri et al., 2015) threshold for phase one households and Beige and Axhausen (2012) who found that mobility stabilises after age 35.

The youngest households were more likely to reside at the city centre, with a large margin separating them from older households. Estimating standard errors and visualising confidence intervals, we were also able to evaluate whether the difference between trend lines was statistically significant or not. We found that the difference between the two youngest household cohorts (15-to-24 and 25-to-34) was also statistically significant. Besides the two youngest groups, the difference between the smoothed location quotient patterns was more difficult to identify, as the standard errors overlap. However, we found that households with a head of household 85 + years or older are statistically more likely to reside in the city centre(s) than households whose householders are between 45 to 74. We found the middle-aged households most likely and the youngest households least likely to reside in the suburbs (starting from immediately after the smoothed trend lines converged). As we predicted in our model, because of overlapping confidence intervals, we found it difficult to distinguish single age groups older that 35 years in the suburbs. At the metropolitan fringe, we observed an almost exact sorting effect as our model suggested, where the probability of residence peaked for older households (householders over 65) and younger households were least likely to reside.

By breaking the age of householders into eight groups, the CLM shows the household sorting effect at a higher resolution than the phasic model. Our analysis showed that age groups exhibit varying behaviours at different distances from city centre(s). For example, using the linear regression lines we were able to identify four clusters of age groups at the city centre(s) and five clusters at the

metropolitan fringe. These clusters justify the phasic model's threshold to identify phase one householders between the ages of 15 and 34, but also showed that instead of age 60 (which was used as a threshold for phase three in the phasic model), housing consumption behaviors show distinguishable patterns after age 65. We also found that middle-aged households (between 35 and 64) can form two to three clusters, depending on the distance from city centre(s) and approach (i.e. recognising patterns with smoothing or linear regression).

The model that we proposed in this paper offers a housing-based solution to population sorting across metropolitan areas. It provides a simple yet functional answer to approximate households' residential locations based on their housing consumption and regional land use patterns. According to our model and our results, young households are the most likely cohorts to reside closer to the city centre(s). This could be due to their low housing consumption/high mobility and the fact that central city locations in the USA often offer lower levels of housing services. In addition, city centre(s) provide more non-residential land uses and other attractive features that also play a role in attracting younger households (Clark and Coulter, 2015). The same rationale could be used to estimate residential locations for low-income or minority households: because they have a lower housing consumption, they are more likely to live closer to city centre(s).

Our model showed that in the USA, the middle-aged households are more likely to live in suburbs. Presence of children plays as a strong motive for mobility stabilisation in more children-friendly neighbourhoods (Beige and Axhausen, 2012; Clark and Morrison, 2012; Coulter and van Ham, 2013; Hedman et al., 2011). For the American middle-aged households suburbs are where they can respond to their housing

needs (e.g. space requirements, etc.), while still affording them, and having access to satisfactory schools. In other words, the suburban development pattern of US metropolitan areas coerces growing households to the suburbs as their only option to live.

Our findings also illustrated complexities in mobility and location choice behaviours of the old. We built most of our assumptions for older households based on the concept of downsizing, which is not a well-studied phenomenon in housing studies (Judd et al., 2014). Downsizing may mean different things to different households. In general, downsizing is perceived as an attempt to establish a new equilibrium between dwelling size and the number of occupants or to improve the quality of the housing unit (Gobillon and Wolff, 2011). For example, downsizing has been observed in the forms of a change from a single family to an apartment (Abramsson and Andersson, 2012) or to a smaller dwelling (Banks et al., 2007). Nevertheless, census data does not provide sufficient data to understand the residential relocation of households as they age – as it is often triggered by involvement of multiple factors that are not collected by the US Bureau of the Census (Judd et al., 2014). Further work and more inclusive data are needed to develop more accurate models of residential distribution patterns for the older population.

The generational sortings we described in this paper have implications for the housing market and policy. In many cities, real estate developers are already incorporating generational cohort-based strategies into their development investment plans. For example, in Melbourne, Australia, high amenity, inner-city high rise blocks are being developed for young adults (Wulff et al., 2010). Cities and local governments can also benefit from understanding the cohort-based model we described in this paper to achieve

and enhance their housing and sustainable development policies. For instance, to promote compact growth, development and zoning policies can focus on providing housing needs of the middle-aged households in city centres – since the middle-aged households are the ones that comprise most of the suburban population.

Reproducing this work

This research has been conducted in a full reproducible manner. 'Reproducibility' in this context means that the empirical results presented in this work can be fully and exactly recreated by other parties using the data and workflow documentation which we have made publicly available. Reproducibility is necessary for the verification of empirical work and, as a result, has been defined as a 'cornerstone' (Mccullough, 2009) or a 'fundamental tenet' (Crick et al., 2014) of good scientific practice. Additionally, research conducted in a reproducible manner has been shown to contain fewer errors (Camfield and Palmer-Jones, 2013), completed more efficiently (Donoho, 2010) and garner more citations (King, 1995).

All code used to create the data and analysis in this paper is available at https://github.com/andykrause/hhLocation. The documentation on this site will provide complete instructions for downloading and executing the necessary code to reproduce our analysis. The raw data (census geographic data and SF1 data) for this analysis are many gigabytes in size. The code will download, extract and clean these data. Users wishing to skip the time-consuming data download and cleaning process may download the cleaned set of data from Harvard's DataVerse repository – available at https://dataverse.harvard.edu/dataverse/repHHLoc. Users are encouraged to use our cleaned data for related research, provided the data are cited.

The model we propose here is also reproducible in other contexts. However, when thinking about external validity of the CLM, one should refer the two basic assumptions that this model was developed upon. Our first assumption that housing career increases over a household's life span should hold globally. Yet, our second assumption about how housing services are being offered across metropolitan regions is US-based and in order for this model to work in other context, this assumption needs to be modified accordingly. Once this assumption is modified, the model in Figure 1 can be reproduced in intersection of the two assumptions and we expect its geometric form to be different for different regions in the world. Future work can reproduce this model in other parts of the world, add more detailed age intervals, or incorporate other proxies for housing career change over time.

Notes

1. To simplify the model, we consider housing provision to refer to structural utility-generating assets such as home size, number of bathrooms, lot size, etc. Location specific amenities – which vary in utility depending on household preference – are ignored.
2. http://www.census.gov/prod/cen2010/doc/sf1.pdf.
3. https://dataverse.harvard.edu/dataverse/repHLoc.
4. Jittering is a function that adds an amount of random noise to the data in order to break ties, and is often used to avoid overplotting.

References

Abramsson M and Andersson EK (2012) Residential mobility patterns of elderly – Leaving the house for an apartment. *Housing Studies* 27(5): 582–604.

Abu-Lughod JL and Foley MM (1960) Consumer strategies. In: Foote NN, et al. (eds) *Housing Choices and Housing Constraints*. New York: McGraw-Hill, pp. 71–271.

Bailey AJ (2008) Population geography: Life-course matters. *Progress in Human Geography* 33(3): 407–418.

Banks J, Blundell R, Oldfield Z, et al. (2007) *Housing price volatility and downsizing in later life*. National Bureau of Economic Research Working Paper Series, No. 13496. Available at: http://www.nber.org/papers/w13496\n http://www.nber.org/papers/w13496.pdf.

Beige S and Axhausen KW (2012) Interdependencies between turning points in life and long-term mobility decisions. *Transportation* 39(4): 857–872.

Boehm TP and Ihlanfeld KR (1986) Residential mobility and neighborhood quality. *Journal of Regional Science* 26: 411–424.

Brown LA and Moore EG (1970) The intra-urban migration process: A perspective. *Geografiska Annaler. Series B* 52(1): 1–13.

Camfield L and Palmer-Jones R (2013) Three 'Rs' of econometrics: Repetition, reproduction and replication. *Journal of Development Studies* 49(12): 1607–1614.

Cervero R and Duncan M (2006) Which reduces vehicle travel more: Jobs–housing balance or retail–housing mixing? *Journal of the American Planning Association* 72: 475–490.

Clark WAV (2013) Life course events and residential change: Unpacking age effects on the probability of moving. *Journal of Population Research* 30(4): 319–334.

Clark WAV and Coulter R (2015) Who wants to move? The role of neighbourhood change. *Environment and Planning A* 47(12): 2683–2709.

Clark WAV and Dieleman FM (1996) *Households and Housing: Choice and Outcomes in the Housing Market*. New Brunswick, NJ: Center for Urban Policy Research, Rutgers.

Clark WAV and Huang Y (2003) The life course and residential mobility in British housing markets. *Environment and Planning A* 35: 323–339.

Clark WAV and Maas R (2015) Spatial mobility and opportunity in Australia: Residential selection and neighbourhood connections. *Urban Studies* 53(6): 1317–1331.

Clark WAV and Morrison PS (2012) Socio-spatial mobility and residential sorting: Evidence from a large-scale Survey. *Urban Studies* 49(15): 3253–3270.

Clark WAV and Onaka J (1983) Life cycle and housing adjustment as explanations of residential mobility. *Urban Studies* 20: 47–57.

Clark WAV and Withers SD (2007) Family migration and mobility sequences in the United States. *Demographic Research* 17(20): 591–622. Available at: http://www.demographic-research.org/volumes/vol17/20/ (accessed 4 January 2014).

Clark WAV, Deurloo MC and Dieleman FM (2002) Housing consumption and residential crowding in US housing market. *Journal of Urban Affairs* 22(1): 49–63.

Clark WAV, Deurloo M and Dieleman F (2006) Residential mobility and neighbourhood outcomes. *Housing Studies* 21(3): 323–342.

Clark WAV, van Ham M and Coulter R (2014) Spatial mobility and social outcomes. *Journal of Housing and the Built Environment* 29: 699–727.

Coulter R and van Ham M (2013) Following people through time: An analysis of individual residential mobility biographies. *Housing Studies* 28(7): 1037–1055.

Crick T, Hall BA and Ishtiaq S (2014) "Can I implement your algorithm?": A model for reproducible research software. *arXiv Preprint arXiv:1407.5981*.

de Groot C, Mulder C, Das M, et al. (2011) Life events and the gap between intention to move and actual mobility. *Environment and Planning A* 43: 48–66.

Denton NA and Massey DS (1991) Patterns of neighborhood transition in a multiethnic world: US metropolitan areas, 1970–1980. *Demography* 28(1): 41–63.

Dieleman FM (2001) Modelling residential mobility; a review of recent trends in research. *Journal of Housing and the Built Environment* 16: 249–265.

Dieleman FM, Clark WAV and Deurloo MC (2000) The geography of residential

turnover in twenty-seven large US metropolitan housing markets, 1985–95. *Urban Studies* 37: 223–245.

Donoho DL (2010) An invitation to reproducible computational research. *Biostatistics* 11(3): 385–388.

Durlauf SN (2004) Neighborhood effects. In: Henderson JV and Thisse J-F (eds) *Handbook of Regional and Urban Economics*. Amsterdam: Elsevier, pp. 2147–2234.

Estiri H, Krause A and Heris MP (2015) 'Phasic' metropolitan settlers: A phase-based model for the distribution of households in US metropolitan regions. *Urban Geography* 36(5): 777–794.

Ewing R and Rong F (2008) The impact of urban form on U.S. residential energy use. *Housing Policy Debate* 19(1): 1–30.

Frey WH (1978) Mover's life-cycle stage and choice of destination neighborhood: Implications for urban social structure. In: Clark WAV and Moore EG (eds) *Population Mobility and Residential Change*. Evanston, IL: Northwestern University, pp. 99–148.

Geist C and Mcmanus PA (2008) Geographical mobility over the life course: Motivations and implications. *Population, Space And Place* 14: 283–303.

Gober P, Larson KL, Quay R, et al. (2013) Why land planners and water managers don't talk to one another and why they should! *Society & Natural Resources* 26(3): 356–364.

Gobillon L and Wolff F-C (2011) Housing and location choices of retiring households: Evidence from France. *Urban Studies* 48(2): 331–347.

Hedman L, van Ham M and Manley D (2011) Neighbourhood choice and neighbourhood reproduction. *Environment and Planning A* 43(6): 1381–1399.

Hunt B and Balachandran B (2015) Black: White disparities in lung cancer mortality in the 50 largest cities in the United States. *Cancer Epidemiology* 39(6): 908–916.

Ihrke D (2014) *Reason for Moving: 2012 to 2013*. Available at: http://www.census.gov/prod/2014pubs/p20–574.pdf.

Ioannides YM and Zabel JE (2008) Interactions, neighborhood selection and housing demand. *Journal of Urban Economics* 63: 229–252.

Judd B, Liu E, Easthope H, et al. (2014) Downsizing amongst older Australians. *AHURI Final Report* 214: 1–197.

King G (1995) Replication, replication. *PS: Political Science and Politics* 28(3): 443–499.

King M, Steven Ruggles J, Alexander T, et al. (2010) Integrated public use microdata series, current population survey: Version 3.0. (Machine-readable database). Available at: https://cps.ipums.org/cps.

Kulu H and Milewski N (2007) Family change and migration in the life course. *Demographic Research* 17: 567–590. Available at: http://www.demographic-research.org/volumes/-vol17/19/ (accessed 1 December 2013).

Landale NS and Guest AM (1985) Constraints, satisfaction and residential mobility: Speare's model reconsidered. *Demography* 22: 199–222.

Lang RE (2002) Open spaces, bounded places: Does the American West's arid landscape yield dense metropolitan growth? *Housing Policy Debate* 13(4): 755–778.

Lee BA, Oropesa RS and Kanan JW (1994) Neighborhood context and residential mobility. *Demography* 31: 249–270.

Lichter DT, Parisi D and Taquino MC (2015) Toward a new macro-segregation? Decomposing segregation within and between metropolitan cities and suburbs. *American Sociological Review* 80(4): 843–873.

Mccullough BD (2009) Open access economics journals and the market for reproducible economic research. *Economic Analysis & Policy* 39(1): 117–126.

Markusen A and Schrock G (2006) The distinctive city: Divergent patterns in growth, hierarchy and specialisation. *Urban Studies* 43(8): 1301–1323.

Moos M (2015) From gentrification to youthification? The increasing importance of young age in delineating high-density living. *Urban Studies* 54(13): 2903–2920.

Nechyba TJ and Walsh RP (2004) Urban sprawl. *Journal of Economic Perspectives* 18: 177–200.

Quigley JM and Weinberg DH (1977) Intra-urban residential mobility: A review and synthesis. *International Regional Science Review* 2: 41–66.

Simmons JW (1968) Changing residence in the city: A review of intraurban mobility. *Geographical Review* 58: 622–651.

Sivak M (2008) Where to live in the United States: Combined energy demand for heating and cooling in the 50 largest metropolitan areas. *Cities* 25(6): 396–398.

Speare A (1970) Home ownership, life cycle stage, and residential mobility. *Demography* 7: 449–458.

Speare A (1974) Residential satisfaction as an intervening variable in residential mobility. *Demography* 11: 173–188.

van der Vlist AJ, Gorter C, Nijkamp P, et al. (2002) Residential mobility and local housing-market differences. *Environment and Planning A* 34: 1147–1164.

van Ham M (2012) Housing behaviour. In: Clapham DF, Clark WAV and Gibb K (eds) *The SAGE Handbook of Housing Studies.* London: SAGE Publications, pp. 47–65.

Waddell P, Bhat C, Eluru N, et al. (2007) Modeling interdependence in household residence and workplace choices. *Transportation Research Record: Journal of the Transportation Research Board* 2003(11): 84–92.

Winstanley A, Thorns DC and Perkins HC (2002) Moving house, creating home: Exploring residential mobility. *Housing Studies* 17(6): 813–832.

Wulff M, Champion A and Lobo M (2010) Household diversity and migration in mid-life: Understanding residential mobility among 45–64 year olds in Melbourne, Australia. *Population, Space and Place* 16(4): 307–321.

Appendix 1. List of the 50 most populated CBSAs in the US according to the 2010 US Census.

CBSA Name	States	2010 Population
New York-Northern New Jersey-Long Island	NY-NJ-PA	18,897,109
Los Angeles-Long Beach-Santa Ana	CA	12,828,837
Chicago-Joliet-Naperville	IL-IN-WI	9,461,105
Dallas-Fort Worth-Arlington	TX	6,371,773
Philadelphia-Camden-Wilmington	PA-NJ-DE-MD	5,965,343
Houston-Sugar Land-Baytown	TX	5,946,800
Washington-Arlington-Alexandria	DC-VA-MD-WV	5,582,170
Miami-Fort Lauderdale-Pompano Beach	FL	5,564,635
Atlanta-Sandy Springs-Marietta	GA	5,268,860
Boston-Cambridge-Quincy	MA-NH	4,552,402
San Francisco-Oakland-Fremont	CA	4,335,391
Detroit-Warren-Livonia	MI	4,296,250
Riverside-San Bernardino-Ontario	CA	4,224,851
Phoenix-Mesa-Glendale	AZ	4,192,887
Seattle-Tacoma-Bellevue	WA	3,439,809
Minneapolis-St. Paul-Bloomington	MN-WI	3,279,833
San Diego-Carlsbad-San Marcos	CA	3,095,313
St. Louis	MO-IL	2,812,896
Tampa-St. Petersburg-Clearwater	FL	2,783,243
Baltimore-Towson	MD	2,710,489
Denver-Aurora-Broomfield	CO	2,543,482
Pittsburgh	PA	2,356,285
Portland-Vancouver-Hillsboro	OR-WA	2,226,009
Sacramento- -Arden-Arcade- -Roseville	CA	2,149,127
San Antonio-New Braunfels	TX	2,142,508
Orlando-Kissimmee-Sanford	FL	2,134,411
Cincinnati-Middletown	OH-KY-IN	2,130,151
Cleveland-Elyria-Mentor	OH	2,077,240

Appendix 1. Continued

CBSA Name	States	2010 Population
Kansas City	MO-KS	2,035,334
Las Vegas-Paradise	NV	1,951,269
San Jose-Sunnyvale-Santa Clara	CA	1,836,911
Columbus	OH	1,836,536
Charlotte-Gastonia-Rock Hill	NC-SC	1,758,038
Indianapolis-Carmel	IN	1,756,241
Austin-Round Rock-San Marcos	TX	1,716,289
Virginia Beach-Norfolk-Newport News	VA-NC	1,671,683
Providence-New Bedford-Fall River	RI-MA	1,600,852
Nashville-Davidson--Murfreesboro--Franklin	TN	1,589,934
Milwaukee-Waukesha-West Allis	WI	1,555,908
Jacksonville	FL	1,345,596
Memphis	TN-MS-AR	1,316,100
Louisville/Jefferson County	KY-IN	1,283,566
Richmond	VA	1,258,251
Oklahoma City	OK	1,252,987
Hartford-West Hartford-East Hartford	CT	1,212,381
New Orleans-Metairie-Kenner	LA	1,167,764
Buffalo-Niagara Falls	NY	1,135,509
Raleigh-Cary	NC	1,130,490
Birmingham-Hoover	AL	1,128,047
Salt Lake City	UT	1,124,197