Agent-based Modelling of Housing Choice in the EASEL Regeneration District

René Janelle Jordan

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
School of Geography

September 2011

Chapter 4 is based on the joint publication:

Jordan, R., Birkin, M., and Evans, A.J. (2011). Agent-based Simulation Modelling of Housing Choice and Urban Regeneration Policy. In: Bosse, T., Geller, A. and Jonker, C. (eds.), Multi-Agent-Based Simulation XI. Springer: Berlin.

This work appears as published in **Appendix H** for reference. I declare that the research for this publication was solely my own work and that I am the lead author. The contribution of the other named authors, Mark Birkin and Andrew Evans, were purely editorial and advisory.

Acknowledgements

I wish to take this opportunity to acknowledge all those who have made my PhD journey an absolutely unforgettable experience. Many thanks to Professor Mark Birkin and Dr Andrew Evans, for your guidance, support and genuine belief in my ability. Also to the Centre for Spatial Analysis and Policy (CSAP) for sponsoring simulation-based research such as this. Doctors Paul Norman, Deborah Phillips, Rachel Unsworth and Stuart Hodkinson of the School of Geography, University of Leeds, thank you for your guidance at various stages of this research project.

Many thanks to Maggie Gjessing (Senior Project Manager East Leeds, Leeds City Council) for taking the time to discuss the regeneration plans for the EASEL district. Your work in this field is admirable, thank you for the engaging discussions on policy and planning as it relates to the EASEL district and the housing market more generally.

To my friends both near and far, thank you for your support; surprise calls, texts, emails, lunch, coffee breaks and trips outside of Leeds. I cannot forget my family who have always been on hand to track progress and listen to the latest developments, your support is invaluable. Mummy and Daddy, this is dedicated to you, thank you for laying the foundation.

Abstract

Housing choice behaviour is made up of a wide array of dynamic individual level behaviour which can be recreated and explored using agent-based modelling (ABM). In the UK, there has been a renewed focus on urban regeneration policies over the past decade by the national government. Such policies seek to deal with the problems of deprivation in communities segregated by socioeconomic status. Using the case study district of East and South Easel Leeds (EASEL), an area known to have pockets of disadvantaged communities, the impacts of regeneration policy will be explored. A computer simulation of residential mobility and regeneration policy is created, in this way, the potential outcomes of regeneration schemes are explored.

Keywords: residential mobility, housing choice, agent-based modelling, regeneration

Published Work

The following is a list of references for peer-reviewed work, book chapter writing and conference presentations.

Peer-reviewed publications

Jordan, R., Birkin, M., Evans, A. (2011): 'Agent-based Simulation Modelling of Housing Choice and Urban Regeneration Policy'. In: Bosse, T., Geller, A. and Jonker, C. (eds.), Multi-Agent-Based Simulation XI. Springer: Berlin.

Book Chapter

Jordan, R., Birkin, M., Evans, A.: 'ABM: Housing Choice, Residential Mobility and Urban Regeneration Policy'. In: Batty, M., Heppenstall, A. and Crooks A. (eds.), Spatial Agent-based Models: Principles, Concepts and Applications. In press.

Presentations

A Social Simulation of Housing Choice and Urban Regeneration Policy. Paper presented at the Regional Science Association International (British and Irish Section) Conference, Glasgow, Scotland. 25-27 August, 2010.

Agent-based Simulation Modelling of Housing Choice and Urban Regeneration Policy, Paper presented at Multi-Agent-Based Simulation Conference, Toronto, Canada, 10-14 May, 2010.

A Social Simulation of Housing Choice, Paper presented at the Housing Studies Conference, University of York, UK, 15 April, 2010.

A Model of Household Preferences and Housing Choice. Paper presented at the Annual Conference of the International Journal of Neighbourhood Renewal, London, UK, 19 November 2009.

A Model of Household Preferences and Housing Choice. Paper presented at the Association of American Geographers, Las Vegas, USA, 22-27 March 2009

Regeneration, Preferences and the Challenge to Mixed Communities. Paper presented at the Royal Geographical Society Annual International Conference, London, UK, 27-29 August 2008.

Large-number Individual-level Modelling of Society: Regeneration and the UK Housing Market. Paper presented at the NCeSS Fourth International Conference on e-Social Science, Manchester, UK, 18-20 June 2008.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introducing the Housing Policy Question

1.1 Introduction

Urban communities are not formed by chance; they are a complex combination of social worlds (TIMMS, D. W.G., 1971). Housing policy has the potential to shape the future of urban communities by bringing about changes in the physical and social landscape. It is one of the instruments used by government to influence the structure of the housing sector and includes as a part of its remit a welfare component which attempts to provide accommodation for and/or improve the dwelling conditions of those unable to provide suitable homes for themselves. Thus deprivation and the state of the poor have played key roles in the development of some housing policies. In recent times, the repeated build and fix housing policy methodologies engaged in by government administrations of yesteryear have turned into an attitude of enabling in an effort to improve the lives of the disadvantaged. These policies have been presented under the umbrella of urban regeneration.

As defined by Bramley *et al.* (2004) regeneration is the process of recovering and renewing lost vitality in the physical and social landscape of a community. Regeneration policy is a combination of projects or schemes. These may include adult education programmes with crèche facilities for young mothers, the provision of additional green spaces such as parks and play areas for children, mixed tenure housing developments as well as retrofitting residential dwellings in need of repair. Such improvements can make a community more attractive to investors and with new businesses established more jobs become available over time for those within the regenerated community (TRUEMAN, M. *et al*., 2004). In this context, physical and social development complement each other; regeneration projects target community development in a holistic way, equipping households with the tools necessary to improve life chances and better support themselves through improved access to employment.

One prime example of a regenerated community is the city of Liverpool. With extensive improvements and development of the urban area, Liverpool gained recognition in 2008 after being named the European Capital of Culture. Hull (2000) argues however, that despite the physical changes in the urban mosaic of most regenerated cities, urban regeneration policy does not result in a narrowing of the gap between the disadvantaged and the rest of society. Instead, urban regeneration is often linked to the process of gentrification; a term used to describe the widening of the gap between high income and low income social classes and the slow exclusion of the latter from regenerated areas (**Section 2.4.3**). Based on these contrasting reports, the potential success of urban regeneration policies may be questioned.

Though there is evidence in the qualitative literature outlining the impact of urban regeneration policy (**Section 2.4.3**), there is a paucity of quantitative literature on this subject area. Quantitative analysis can be used to further examine hypotheses made in the qualitative literature through statistical analysis and/or computational models and simulations. Simulation allows for the use of actual real world data and is used to understand phenomena in the real world as well as forecast outcomes in the future. As a result, computational modelling will be used to analyse the effects of urban regeneration schemes proposed for the case study area in the east and south east districts of Leeds, UK. More specifically, the technique of agent-based modelling (ABM) will be used to build a model representative of residential communities in Britain. Data from the East and South East Leeds (EASEL) regeneration district will be used to assess the possible outcomes of regeneration projects to be implemented in this community.

In this brief introduction, the research question, aims, objectives and hypotheses are presented. The case study area of EASEL will be described as well as the proposed regeneration plans for the district. The introduction is concluded with summaries of subsequent chapters.

1.2 Research Question, Hypothesis, Aims and Objectives

What insights can ABM give on the impacts of regeneration policy is the research question posed. The general hypothesis suggest that regeneration projects can lead to socially mixed communities. This hypothesis is further discussed in **Section 2.4.3**.

The aims and objectives are as follows:

Aims

1. To build an agent-based model to represent the relationship between economic and social drivers by investigating residential mobility behaviours.
2. To create features within the agent-based model to simulate the implementation process of regeneration projects.
3. To forecast the social mix of the population after regeneration in terms of socioeconomic diversity and ethnic diversity.

Objectives

1. To review the literature on urban regeneration and the housing market, and to elicit expert opinion on these subject areas so as to determine the residential mobility behaviours.
2. To build a hybrid computer simulation merging the concepts of microsimulation and ABM in order to recreate residential mobility behaviour and implement regeneration projects proposed for the EASEL area.
3. To use the residential mobility model to assess the change in the socioeconomic and ethnicity mix of the population over a 20 year period commencing in 2001.

1.3 A Description of the EASEL Case Study Area

With a population of over 700,000 residents as at the 2001 UK Census, Leeds is one of the largest metropolitan districts in England (REES, P. *et al.*, 2004). The city is characterised by a booming financial sector and a large student population. Despite this, the city of Leeds is noted to contain some of the most deprived communities in

England, with at least 35000 Leeds residents living in areas rated amongst the 3% most deprived in the country (LEEDS CITY COUNCIL, 2005). These areas include Lincoln Green, Burmantofts, Harehills, Gipton, Seacroft, Halton Moor, Osmondthorpe, Richmond Hill and Cross Green.

As a result of this, Leeds City Council intends to invest in the redevelopment of this urban community. Much of the land around the city is earmarked for redevelopment aimed at improving neighbourhood quality, providing educational opportunities and improving the state of the local economy to reduce unemployment. For example, private homes are to be offered for sale as well as affordable housing for low-income households. Provisions for better access to goods, services and leisure activities are some of the other features of the regeneration plans (LEEDS CITY COUNCIL, 2005).

One area chosen to benefit from this programme of redevelopment is the EASEL district. In general, the EASEL district is home to approximately 78000 people living in 35000 households (LEEDS CITY COUNCIL, 2007a). The district is an area noted to suffer from high levels of deprivation when socioeconomic variables are assess as well as the negative effects of crime, violence and antisocial behaviour. EASEL is also resident to a large number of social housing tenants, several of whom rely on welfare support. Statistical data detailing the socioeconomic variation and demographic makeup of this district are presented in this section. This data has been derived from the EASEL Needs and Aspirations Study as well as the EASEL Area Action Plan while other EASEL specific data related with preferences and housing trends are presented in **Section 2.3**. The EASEL area is made up of the wards of Burmantofts, Harehills, Seacroft, and Richmond Hill as illustrated in **Figure 1.1**. Note that MLSOA is a census geography of approximately 7200 individuals (OFFICE OF NATIONAL STATISTICS, 2011b).

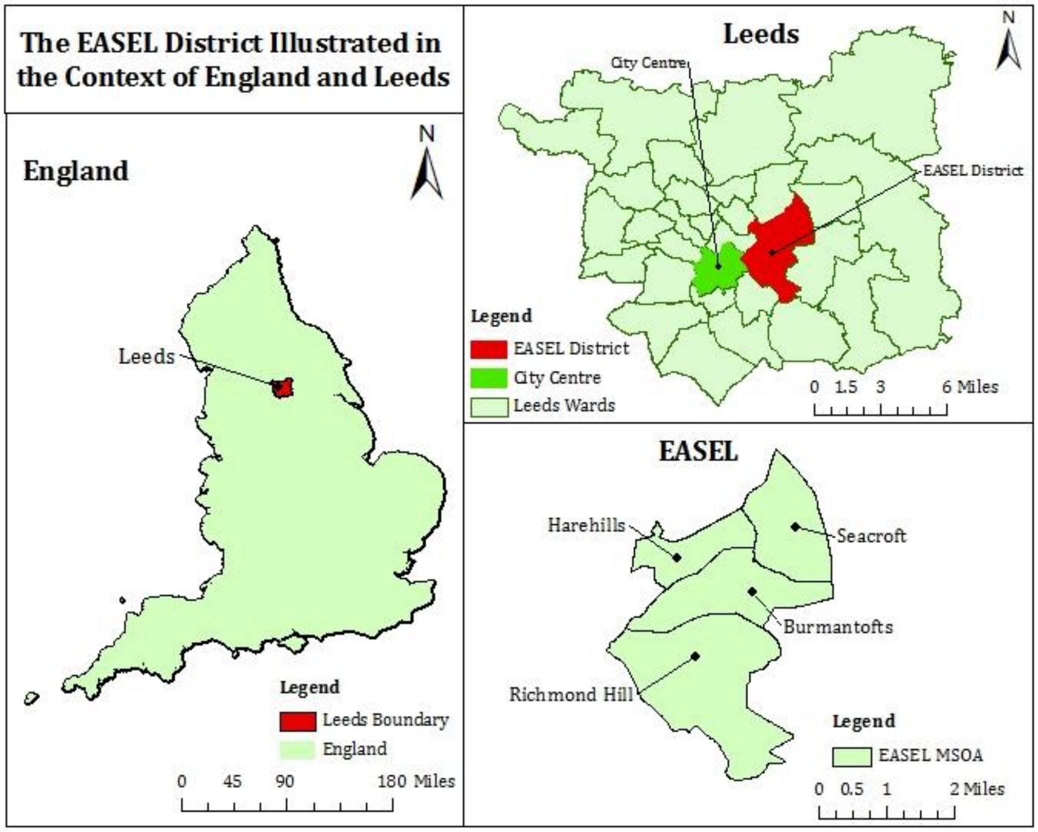


Figure 1.1 The EASEL District Illustrated in the Context of England and Leeds by MLSOA

Using the 2004 Indices of Deprivation, of the 53 MLSOAs of the EASEL district, 45 were noted to fall within the top 10% deprived areas across the nation (**Table 1.1**) (LEEDS CITY COUNCIL, 2007a). This is in stark contrast to the Leeds Metropolitan District also noted in the table.

Area	**Number of MLSOAs**	**Number of MLSOAs in top 10% nationally**	**Rate**
EASEL	53	45	85%
Leeds	476	100	21%

Table 1.1 Number of deprived MLSOAs in the EASEL District versus the Leeds Metropolitan District

As of the 2001 census, 41% of individuals living in this area were noted to be economically inactive; individuals seeking employment (LEEDS CITY COUNCIL, 2007b). A further 5% were unemployed; those individuals not able to seek employment due to long term illness among others (LEEDS CITY COUNCIL, 2007b). In terms of economically active individual, 58% of EASEL residents fell into this category. Within the same time period, 34% of Leeds residents were noted to be economically inactive while a further 3% were unemployed (LEEDS CITY COUNCIL, 2007b). This can be contrasted to the Leeds district figures where unemployment and economically inactive figures are lower than those for the EASEL district while economically active figures are higher. These figures are illustrated in the table below, **Table 1.2**.

Economic Activity	EASEL		Leeds	
All People	**53228**	**100%**	**520479**	**100%**
Economically Inactive	22160	41.63%	177,773	34.16%
Economically Active	31068	58.36%	34706	65.84%
Unemployed people	2854	5.36%	17280	3.32%

Table 1.2 Economic Activity in the EASEL area compared to Economic Activity in Leeds

Reported crime rates as shown in **Table 1.3**, are higher for the EASEL district as compared to the Leeds district when the criminal damages category is considered. For domestic burglary and vehicle crimes, the statistics for the EASEL district and Leeds are considerably comparable when the rates of crime are considered (LEEDS CITY COUNCIL, 2007b).

Crime	EASEL	Rate	Leeds	Rate
All Reported Crime	**15493**	**100%**	**98320**	**100%**
Domestic Burglary	1219	7.87%	7793	7.93%
Vehicle Crime	1879	12.13%	12826	13.05%
Criminal Damage	4280	26.63%	22073	22.45%

Table 1.3 Crime Statistics for the EASEL district compared to the same for the Leeds district. Source: West Yorkshire Police 2005

The area is a mixture of accommodation types including tower-block purpose-built flats, terraces, detached and semi-detached homes, some of which are in poor condition. The latter may be due to the lack of a programme of sustained maintenance over the years. **Figure 1.2** presents a set of images taken around various communities within the EASEL district.

(a) (b) (c)

Figure 1.2 EASEL urban landscape. Scene (a) Terrace housing among tower blocks. Scene (b) Green spaces available around tower block houses. Scene (c) Terrace houses in poor condition, originally built as temporary housing after the Second World War.

Housing tenures range from ownership, private rented and social housing with 41% of EASEL residents living in council housing and 8% in housing provided by Housing Associations. This may be compared to the Leeds district where 20% of the population reside in social housing and 4% rely on housing provided by Housing Associations (LEEDS CITY COUNCIL, 2007b) (**Table 1.4**).

Crime	EASEL	Rate	Leeds	Rate
All Households	**33535**	**100%**	**301623**	**100%**
Owner occupied	12693	37.85%	187645	62.21%
Social Housing: Council	13970	41.66%	63075	20.91%
Housing Association	2683	8%	12990	4.31%
Private Rented	4189	12.49%	37913	12.57%

Table 1.4 Counts of Tenure Types for the EASEL district compared to the same for Leeds adapted from the 2001 census

When accommodation types in the EASEL district are compared to those of the entire Leeds community, there are considerably less detached homes in the EASEL district than the Leeds community, while terrace housing and purpose built flats appear to be more popular in the EASEL district than in the wider Leeds community. The latter may be related to the wave of high-rise homes built by government after the Second World War in an effort to house a large number of households on one housing plant (**Section 2.4.1**). These rates are detailed in **Table 1.5** below.

Households spaces and accommodation type	EASEL	Rate	Leeds	Rate
All Households with residents	**33520**	**93.62%**	**301614**	**96.52%**
Vacant houses	2285	6.38%	10861	3.48%
Detached	1280	3.57%	46108	14.76%
Semi-detached	13557	37.86%	121394	38.85%
Terraced housing	12953	36.18%	87361	26.96%
Purpose built flats	7640	19.40%	44179	14.14%
Flat/Maisonette/Shared house	343	2.9%	13115	4.2%
Temporary Structure	24	0.07%	398	0.13%

Table 1.5 Counts of Households spaces and Accommodation types for the EASEL district compared to the same for Leeds adapted from the 2001 census

In addition, the area is also home to high concentrations of ethnic minority groups segregated from the wider communities. There is evidence to suggest that this type of segregation creates disconnects between minority groups and the rest of the population (GIBSON, J. *et al*., 1999). Historically, the EASEL community is one which has attracted a higher than average number of ethnic minority groups by way of immigration. Overtime and due to the locality of specific amenities such as places of worship, familiar goods and services, the area grew increasingly attractive to these minority groups (STILLWELL, J. and Phillips, D., 2006). The fear of harassment also amplified this clustering. Coupled with the effects of poverty which early migrants may have encountered, ethnic segregation and poverty may appear to be synonymous though Stillwell and Phillips (2006) notes that many minority groups opt to live in supposedly deprived communities because of the social networks and religious institutions available. In general, though the area is largely populated by White British households, the number of households in minority groups is higher for the EASEL community than for the entire Leeds district as illustrated in **Table 1.6** below.

Ethnicity Group	EASEL	Leeds
White	84.41%	91.85
Non-White	15.59%	8.15%

Table 1.6 Percentage of households by Ethnicity Group in the EASEL District and Leeds

In terms of demographic details, the EASEL district is comprised of a larger number of children aged 0-15 than the entire Leeds community though for older age groups, the population counts for each age categorisation is comparable for both the EASEL and Leeds district as shown in **Table 1.7** (LEEDS CITY COUNCIL, 2007a).

Ethnicity Group	EASEL		Leeds	
All age groups	**78702**		**715402**	
0 – 15	20308	25.81%	143091	20%
16 – 19	4283	5.44%	39439	5.51%
20 – 29	10603	13.47%	108981	15.23%
30 – 59	28917	36.74%	281176	39.3%
Over 60	14592	18.54%	142705	20%

Table 1.7 Comparing the counts of individuals by age groups for the EASEL district and the wider Leeds community

Thus, given the high incidences of welfare dependence within low income households, higher than average levels of crime and violence, high levels of unemployment and deprivation across the EASEL community, Leeds City Council believes that the area is in need of change.

1.4 EASEL Regeneration Plan

The EASEL regeneration plan is a housing-led regeneration scheme aimed at reducing the gap between the most disadvantaged and the rest of the community (LEEDS CITY COUNCIL, 2007a). Originally the project was earmarked to begin in 2008, spanning a period of 15-20 years, and governed by details in the EASEL Area Action Plan (AAP). The EASEL AAP is a collection of proposals which broadly aims to create communities in which people willingly choose to live and work. In this way, it is the belief of the Leeds City Council that sustainable communities can be created (LEEDS CITY COUNCIL, 2007a). Note that the EASEL AAP is a dynamic plan which is designed to react to economic and other changes; and for this reason, the plans outlined may be subject to change at any time.

The key objectives of the EASEL AAP, as documented by the Leeds City Council (2007a), focus on improvements in housing quality and housing opportunities by increasing the available housing stock by tenure and type; the provision of job opportunities; improvements to road network links; provision of more opportunities for healthy lifestyles such as green spaces and leisure facilities; improvements in the provision of suitable access to goods and services; improvement in the quality of the natural environment and improvements in the built environment.

Of primary interest to this research project is the role of housing in the regeneration scheme. The council intends to introduce a greater mix of housing tenures in council owned areas by introducing housing for sale and for rent on the private market. It is felt that a greater mix of tenures leads to greater socioeconomic diversity (LEEDS CITY COUNCIL, 2007a). Sustainable communities are suggested by the council to be the result of successful mixed communities. Leeds City Council (2007a) believes that sustainable communities improve the stability of the housing market.

In order to facilitate this, an estimated 7800 new homes are to be built to create these new mixed tenure communities, with at least 20% affordable housing available to social housing tenants interested in private sector housing by 2011 and at least 25% affordable housing between 2012 and 2016 (LEEDS CITY COUNCIL, 2007a). It is expected that several council housing estates will be demolished in a piece-meal fashion to accommodate the new housing. Displaced council tenants are expected to be reallocated to other social housing units while those social housing tenants that can afford it will be encouraged to take advantage of the affordable housing provisions. In this way, social housing tenants are able to join the owner occupier market. It is through initiatives such as these that the council hopes that deprived communities within the EASEL district may be transformed into sustainable communities. It should be noted, however, that due to the economic downturn in the 2008-2009 period and a subsequent change in political governance during the 2010 general elections, some of these plans have been changed. For example, though the original plan indicated that 7800 new homes were to be built, today this number has been significantly reduced. Alterations such as these do not directly affect this project, however, as the likely effects of a variety of proposed projects can still be modelled. A detailed account of the AAP for the EASEL district can be found in the EASEL Area Action Plan 2007 produced by the Leeds City Council (LEEDS CITY COUNCIL, 2007a).

1.5 Chapter Overview

Having introduced the research question, the case study area of EASEL and the proposed regeneration policy for this area, this chapter overview will briefly describe each of the proceeding chapters.

Chapter 2 discusses the themes of residential mobility, housing policy and the impacts of regenerating disadvantaged communities. Much of the extant literature is discussed with a view to understanding the process of residential mobility as well as identifying mobility preferences and behaviours. These behaviours are those that govern the types of households who choose to relocate and the housing choice process undertaken when finding a new home. Housing policies and their effects on the housing market over the last century are discussed. Urban regeneration as a policy is discussed in detail with special emphasis placed on the impact of regeneration programmes and the potential negative impacts identified in the qualitative literature. Though the qualitative literature outlines the possible effects of regeneration projects, there is little documentation on the actual impacts of past regeneration programmes. For this reason, the chapter is concluded with the notion that quantitative analysis is needed in this area to fully analyse the information presented in the qualitative literature.

In **Chapter 3** computer modelling is presented as a quantitative methodology that can be used to examine the research question. The benefits of computer modelling are highlighted. Computer models are abstractions of reality and are used to examine theory and/or existing hypothetical claims. Several types of modelling techniques are presented, namely microsimulation modelling, cellular automata (CA) modelling, spatial interaction modelling and ABM. Examples are used to illustrate the usefulness of each modelling technique, though ABM is focused on as the modelling technique of choice due to its advantages in implementing and manipulating individual level behaviour dynamically. Several agent-based models of residential mobility are presented, comparing and contrasting the types and quantity of dynamic behaviours simulated. A new model, the CHAIRS simulation, is presented as an alternative to the existing agent-based residential mobility models. CHAIRS, an acronym for **C**reating

Housing **A**lternatives **I**n **A** **R**ejuvenated **S**ociety, introduces additional behaviours when compared to existing models of residential mobility.

Chapter 4 is a presentation of the methodological framework for the CHAIRS simulation. The chapter is presented in the context of model inputs, model processes and model outputs. Real world data is inputted into the model. This data includes census data representing households created through the process of microsimulation and spatial data used to represent neighbourhood entities such as houses, output areas (OA), roads and schools. Note that a household is used to represent a group of individuals residing in the same house. The algorithmic details of the model process are then presented. Here the flow of the model is presented highlighting the process of determining if households should move and the process of finding a new home. These decisions are influenced by the behavioural rules implemented in the model, each of which is verified in this chapter. General model outputs are then presented. Of particular interest in this chapter is the list of assumptions made in the creation and execution of the CHAIRS simulation. The discussion highlights the importance of all model assumptions. With the methodological framework presented, the simulation can be executed, paving the way for the model to be calibrated and validated.

Chapter 5 documents the process of calibrating and validating the CHAIRS simulation. The calibration/validation datasets derived from Acxiom's Research Opinion Poll (ROP) are described in detail. Behavioural rules in the model are noted as the parameters to be calibrated. The CHAIRS simulation is executed with several different combinations of these behavioural rules then the calibration/validation process is applied as a test to the results. The validation proceeds by comparing the results of the model with the results of the ROP and compares absolute counts of households in each OA. The behavioural rule set combination generating the least number of errors is then optimised and used to generate the final results of the model. The chapter ends by critiquing the performance of the CHAIRS simulation, providing support to suggest that all results generated are sufficiently credible.

Chapter 6 is a detailed presentation of the model results when a select number of regeneration projects are implemented. There are 3 scenarios presented: scenario 1 outlines the addition of a new mixed tenure housing development, scenario 2 allows for a change in a major road network directly joining the north and south of the EASEL district, scenario 3 is a combination of scenarios 1 and 2. Each of these scenarios is implemented by altering the spatial data inputted in the model. The model is also executed without the implementation of the scenarios; this is called the baseline situation. In this way, the results of the scenarios can be compared to the baseline results so as to compare likely outcomes in the EASEL district when regeneration projects are implemented and when projects are not implemented. All scenarios are run over a simulation time period equivalent to 20 years, that is, 2001 to 2021. Issues such as gentrification and neighbourhood stability are discussed in the context of these results.

Finally, **Chapter 7** is a summary of the project and details the major findings of the CHAIRS simulation. Reflections on this modelling exercise are discussed including the main challenges encountered. Additional research themes arising from this project are briefly identified and examined. These have been suggested as topics of interest for future research.

Chapter 2
Residential Mobility, Housing Policy and the Plan to Regenerate Disadvantaged Communities

2.1 Introduction

The study of residential mobility is one of many interdisciplinary topics which has captured the interest of researchers in both academic and non-academic spheres. Formal research on this subject dates back to Rossi (1955), who explored the reasons why households change location, strongly linking it to changes in the life course and changes in the circumstances of households. Over the years, other work has evolved which further extends what was known about residential mobility. **Table 2.1** is an abridged list which highlights some of this research and the associated researchers.

Residential mobility and...	Authors
Migration	(BROWN, L. A. and Moore, E. G., 1970)
The Life Course and Housing Choice	(MCCARTHY, K., 1976)
Life Cycle and Housing Adjustment	(CLARK, W.A.V. and Onaka, J. L., 1983)
Segregation	(PHILLIPS, D., 1998)
Housing Tenure and Labour Markets	(BÖHEIM, R. and Taylor, M., 1999)
School Mobility	(CROFT, J., 2004)
Location Choice Behaviour	(KIM, J. H. *et al*., 2005)
Neighbourhood Change	(CLARK, W.A.V. *et al*., 2006)
Neighbourhood Quality	(RABE, B. and Taylor, M. P., 2010)

Table 2.1 Extended research themes in the study of residential mobility

Residential mobility is a process initiated by a decision to migrate and follows on with the selection of and relocation to a new home. In general, it is thought that the decision to migrate is the result of a change in circumstance of households which prompts the need to search for a new residence ((ROSSI, P. H., 1955); (DIELEMAN, F., 2001); (RABE, B. and Taylor, M. P., 2010)). In fact, Strassman (2001) asserts that when the characteristics of a house no longer satisfy the characteristics of a household, the household will try move. If no suitable alternative residence can be found, households may reassess their preferences and search for a new house based on the new requirements or they may opt to remain at the current residence for a longer time

(BROWN, L. A. and Moore, E. G., 1970). As suggested by McCarthy (1976), in this case, relocation may be inhibited if there is no change in household income.

The factors which influence households to have a desire to move are many and the subsequent process of finding a new home is driven by household preferences. These preferences are constrained by market forces and other regulatory factors such as housing stock supply (TU, Y. and Goldfinch, J., 1996). When households move, the results of the interaction between preferences and constraints can affect the pattern of settlement across society ((PEACH, C., 1996); (AGUILERA, A. and Ugalde, E., 2007)). For example, if the housing options within wealthy neighbourhoods and working class neighbourhoods are compared, though households may prefer to live in a specific type of house, they are limited to houses they can afford. This is due to the fact that land prices vary according to amenities and features of the land, affording developers the opportunity to build houses in such a way that the house prices are positively correlated to the land value and neighbourhood quality. More practically, young professionals may be contrasted to mature families. Whereas families may be more concerned with the availability of amenities such as green spaces and prefer to avoid built-up urban areas, constrained by their incomes, young professionals may find these urban areas more accessible.

Housing policy creates an environment where exogenous constraints can be regulated to satisfy the overall housing needs of society (HOLMANS, A.E., 1987). In this way, problems such as those related to the inequities between demand and supply can be addressed and the effects of economic policy such as changes in interest rates can be cushioned; as in the case of the rent to ownership intermediate market schemes presently offered by Local Councils. Thus housing policy is able to constrain the market or can open opportunities to reduce the effects of the constraints that limit households. It may be described as reactive; changes in the population profile can be observed, assessments on the overall welfare of communities can be made and responses in the form of housing policy may be enacted in an attempt to improve the welfare of a community. This may lead to improvements in the availability of educational, recreational and healthcare facilities as well as other services. There is therefore a direct link between housing policy and residential mobility (BROWN, L. A. and Moore, E. G., 1970).

Though housing policy is able to regulate activity in the housing market in general, some of these policy interventions are geared toward helping the poor and socially disadvantaged, many of whom rely on the housing provisions of the public sector (MALPASS, P., 1999). UK council housing has been experienced by numerous problems over the years; for example, high unemployment and persistent welfare dependency. It seems apparent that many of the changes made by government were with the aim of getting rid of these problems, but today, challenges in this sector still persist. Recent housing policies, geared towards urban regeneration, aim to significantly reduce the problems experienced on council housing estates. Regeneration policies aim to build sustainable communities by creating mixed-tenure housing developments and by meeting the observed needs of communities in the form of social care (LEEDS CITY COUNCIL, 2005). However, the effectiveness of regeneration plans is yet to be documented. Can urban regeneration policies lead to the evolution of sustainable communities? If the process of residential mobility is considered in the context of past housing policies it is apparent that past housing policies had a role to play in creating some of the problems in this sector; for example, the problem of residualisation discussed in **Section 2.4.1**. With this in mind, a degree of scepticism should be applied when the benefits of regeneration are presented.

In this chapter, the factors which influence residential mobility will be discussed. Specific reference will be made to residential mobility behaviour in the EASEL district which will be contrasted to the residential mobility behaviour identified for the wider Leeds community. A discussion on housing policy follows which highlights the role housing policy has played in shaping the housing market. Though there is not sufficient literature reporting the actual outcomes of regeneration projects, by examining the trends in the housing market as a result of policy interventions, the likely results of regeneration programmes are explored.

2.2 Understanding why households move and where they go

Households and individuals move for many reasons. Traditionally, reasons for moving have been strongly tied to changes in the life course or life cycle (ROSSI, P. H., 1955). However, Clark and Onaka (1983) point out that due to the increasing number of non-traditional households, it is more appropriate to link the decision to move to reasons related to household dissatisfaction rather than the life course. Non-traditional households may include, for example, unrelated individuals living in shared accommodation.

Changes in household size as well as changes in employment status, neighbourhood quality and/or dwelling quality are likely reasons for a disparity between a dwelling and a household occupying it (RABE, B. and Taylor, M. P., 2010). Partnership formation or dissolution, births and deaths, and children leaving the family home are life course events that alter the trajectory of households and cause them to expand or contract ((CLARK, W.A.V. and Huang, Y., 2003); (MULDER, C. H., 1996)). Employment gains or losses may affect what can be afforded and improvements in neighbourhood quality may be sought after (BÖHEIM, R. and Taylor, M., 1999). Also, as a result of the deterioration of dwelling quality, a new residence may be needed. Such changes directly influence the desire for reduced or increased dwelling features – bedrooms, nearby green spaces etc. In the case of employment, relocating to areas where job opportunities are available may be necessary. Triggers such as these highlight the differences between what households want versus what they have, and as such begin the residential mobility process.

2.2.1 Life Course Triggers

Whether described as a life course or a life cycle, McCarthy (1976) suggests that individuals graduate through phases of family creation and dissolution over their life span. These stages are largely driven by the age of the individual. At each stage of the life course, different housing requirements are thought to be needed in order to match the characteristic needs of the household with an appropriate dwelling (DIELEMAN, F., 2001). In general, the life course begins with a single individual, and over time, family units are formed with or without children. Children eventually leave the family home, beginning their own life courses. Finally, parents revert to a stage of childlessness until death brings their own life course to an end.

If leaving the parental home is considered to be the start of the life course, most individuals are generally young in age (CLARK, W.A.V. and Onaka, J. L., 1983). Such persons are thought to be new entrants to the job market and/or seeking higher education. In essence, this is the stage where mature children leave the parental home in pursuit of their own interests. Incomes are generally low due to their lack of seniority and experience (MCCARTHY, K., 1976). As a consequence of this, renting is the tenure of choice and the propensity to move is highest when compared to other life stages. The pursuit of opportunities, wherever they arise, is foremost on the agenda particularly for highly educated individuals (LONG, L., 1998). Apart from rented accommodation, at this stage, small purpose-built flats, close to the city centre are most popular (SOUTH, S. J. and Crowder, K. D., 1997).

As the individual becomes older and more established, the propensity to move, though high, reduces when compared to the previous stage (LONG, L., 1998). Moving to opportunity – from one job to another or leaving institutions of higher learning – these young adults begin forming family units through marriage or partnerships. Often childless, the requirement for green space around the home is initially not as important while rented accommodation is attractive and incomes become more stable (SOUTH, S. J. and Crowder, K. D., 1997). With incidents such as the birth of a child, couples move on to more family type housing – terraces, semidetached and detached homes – which are often outside of the heart of the city centre. More likely than not, the owner-occupier market is chosen once it can be afforded. At this stage, couples often purchase bigger homes than necessary as they anticipate the continued growth of the family (LONG, L. H., 1972). Thus the size of the house is one consideration that does not always fall in line with the present family size. This is also the stage where the quest for good schools becomes important and householders may opt to move to prime catchment areas as children grow from toddlers to elementary school age and then on to the teenage years ((LONG, L. H., 1972); (CROFT, J., 2004)).

The next stage relates to more settled, mature families. With established jobs and well integrated in their community, again the propensity to move is reduced (LONG, L., 1998). Often living further away from the city and with more established incomes, these families opt for owner occupation. Mulder and Hooimeijer (1999) link the stability at this stage to the tenure type and size; that is, owners are generally less likely to move as the costs of moving can be prohibitive (LEE, B. A. *et al*., 1994).

As children move out of the parental home, the final stage of the life course begins. Parents are left with houses which may be larger than necessary but they may opt to continue to live in these homes to reduce the hassle and costs of moving and because of their strong ties to the community (LEE, B. A. *et al*., 1994). As the years progress, parents enter into retirement, an event which causes a significant drop in income and at times make the maintenance of these larger dwellings difficult. Coupled with the consequences of old age, and with untimely events such as the death of a spouse or partner, smaller homes within relatively close proximity to goods and essential services may be sought after (BANKS, J. *et al*., 2011). Thus, homes in the rental market may also become more attractive.

Evident in the transitional phases of the life span is the progression from the rental market to home ownership. Kemp and Keoghan (2001) examine this progression and liken it to a ladder, where the top rung of the ladder represents home ownership. The ladder illustrates how movement from one tenure type to another is based on a hierarchy of tenures; households are thought to step on to the housing ladder by entering the rental market though ultimately aspiring to become home owners. Households then move from the private rental market to social housing or home ownership. Here social housing is used as a stepping stone to home ownership. The purpose of social housing in the context of UK history has varied considerably over time. Present literature suggests that government may be adopting this type of intermediate market (MEGBOLUGBE, I. F. and Linneman, P. D., 1993) as is discussed in **Section 2.6**.

2.2.2 Other Triggers

The view of the life course as one wholly comprising of family oriented living has limitations when used to describe households in the present day. With young people opting to marry later in life, cohabiting adults and house sharing fast becoming other models of household structure, the traditional view of the life cycle is broken. In the case of these non-traditional household structures, individuals continue to move to improvement in search of better jobs and quality of life unless set back by other life events (DIELEMAN, F., 2001). Thus, linking the decision to move to life cycle events alone is not sufficient. It is more appropriate to link such a decision to overall change in individual circumstances and/or other life events (CLARK, W.A.V. and Onaka, J. L., 1983).

Ultimately, the decision to move and the choice of a new home is constrained by the individual's financial budget (DIELEMAN, F., 2001). Individuals may compromise on household features and other amenities based on this constraint. For example, an individual with limited disposable income may seek low cost housing. The consequence of this may be a compromise on neighbourhood quality; for example, areas where problems of crime and disorder may at times be problematic may be selected. Therefore, compromises on dwelling quality may be made as a means of living within the available budget though household features such as dwelling age and overall quality tend to be subjective; depending on what is aesthetically pleasing to the household (TU, Y. and Goldfinch, J., 1996). In general, the tenure chosen is dependent on the life course stage in collaboration with the income; a settled individual who can afford to purchase a house is more likely to do so than a young professional in search of a better job.

Finding a better job often initiates the mobility process. This is especially true if the major breadwinner of the household is the one to find the job. If the job is in a different region then a long distance move is necessary. If the job is within the local vicinity and offers an increase in earnings then this may trigger a move to a better house (DIXON, S., 2003). The lost of a job may also necessitate a move as earnings are significantly decreased, while unemployed individuals may be willing to move to other areas where job opportunities are thought to be more readily available (BÖHEIM, R. and Taylor, M., 1999).

Local moves are often short distance migrations within a distance of five miles away from the previous house, as individuals tend to move to areas with which they are familiar (RABE, B. and Taylor, M. P., 2010). This also reduces disruptions to a minimum. That is, short distance moves retain reasonable access to transport routes for travelling to work, schools if children are involved, doctors surgeries, shops and other services.

Another factor to be considered is the location of the new home in relation to good schools. Even before school age, some households think about the type of schools they wish their children to attend and try to find houses in such areas (CROFT, J., 2004).

The aim is to find a house within a reasonable catchment distance and where the school is thought to be achieving good results.

In the case of ethnic minority groups, though life course events may occur, altering the size of the home required, such households tend to find housing in areas where there are others of the same or similar ethnic group. This is particularly true for Leeds where the fear of harassment or actual incidences of racial prejudice are factors which trigger the decision to move (PHILLIPS, D *et al.*, 2002). Again, this leads to a clustering of similar households of the same ethnic background (PHILLIPS, D., 1998). Phillips *et al.* (2002) further highlights the behaviour of South Asians in the EASEL area, noting that these groups do not display their wealth in material possessions. Instead, such groups appear to concentrate on the overall neighbourhood make up, opting to live in neighbourhoods where there are others of the same ethnic group. As a consequence of this, neighbourhood quality may be sacrificed (PEACH, C., 1996). Neighbourhoods such as these become saturated and often are segregated from the rest of society. Such communities are close knit; familiar goods and necessary services are easily accessible within the community circle.

2.2.3 The Case of the Council tenant

Though council tenants are faced with the same type of life course changes, the factors that influence them are slightly different from those on the regular market. As they transition from one stage of their life course to another and housing requirements change, those reliant on council housing are limited to the available housing stock within the Local Authority or may be forced to move to another Local Authority to find housing. In addition to this, forced moves can be triggered by the council. These moves may be a result of refurbishment plans or the total demolition of blocks of council owned housing, as is the case with some regeneration schemes. Voluntary moves on the part of the council tenant may at times occur. Such moves may be triggered by dissatisfaction with housing and/or neighbourhood quality; high crime statistics, increases in antisocial behaviour etc. In general, council tenants are limited to where council housing is located.

Events along the life course can trigger the decision to move. Where households go is in part influenced by this hierarchy of tenures which is evident in the way the housing market works. It is important to realise however, that though life course events trigger movement in the housing market, the factor which appears to have the greatest potential to limit or hinder progression up the housing ladder is the earning capacity of the household (MCCARTHY, K., 1976). Households therefore choose tenure types according to what they can afford. As a result, some households may not progress to home ownership though they may wish to move. This is one example of where the relationship between housing policy and residential mobility can be seen. With home ownership being the ultimate goal on the housing ladder for many, one of the aims of housing-led regeneration projects is to give households stuck in the private or social housing market an opportunity to progress to home ownership by the creation of affordable home schemes.

2.2.4 Exogenous factors

Residential mobility is affected by environmental factors which bear down on the housing market. Most notably the economic recession; which has significantly affected house prices and the housing market in general. Some properties may become increasingly more attractive as market and institutional factors change. These factors include mortgage rates, rates of return of housing investments and government intervention in housing supply (PAWSON, H. and Bramley, G., 2004). A household's choice of dwelling is constrained by the demand for and the requisite supply of housing as well as the availability and accessibility of financial instruments such as mortgages. This ability to access a mortgage is dependent on the household income. However, these determinants are governed by activity at the local, national and international level.

Dieleman (2001) suggests that the fluctuation of the economy, inflation, mortgage rates and demographic change play a role in influencing decisions. This is the case as changes in interest rates or fluctuations in the economy have direct effects on house prices. In a similar way, this economic activity can encourage or discourage house construction that in turn affects supply and, through this, demand. As mortgage and other interest rates change, lending agencies may alter the requirements for these credit facilities. For example, an increase on the initial deposit needed to purchase a house may be a deterrent to a household unable to raise the deposit. A household experiencing positive equity may choose to move to make a profit on their house. Similarly, a household experiencing negative equity may be forced to stay in the present house even if there is a desire to move ((BÖHEIM, R. and Taylor, M., 1999); (HENLEY, A., 1998)).

Estate agents also play a role in matching households to neighbourhoods according to the household's individual characteristics. By withholding information on housing opportunities and marketing areas thought to be more suited to the household, estate agents are able to channel households into certain types of houses in specific neighbourhoods (HICKMAN, P. *et al*., 2007).

Households are also affected by vacancy chains where a move may require a period of waiting for each household (FERRARI, E., 2011). In the owner occupier market, households may have to wait until a buyer for the present residence is found and in turn they may have to wait until the new house becomes vacant. In some cases a move to rental accommodation is necessitated if the buyer for the present house is found before the new house is found or becomes unoccupied. In the latter case, owners of the new house may also be waiting until their intended new house becomes vacant (FORREST, R and Murie, A, 1994). To some extent, a similar phenomenon exists in the private and public rental markets.

2.2.5 Effects on settlement patterns

Using the life course alone one can build a picture of how the population is ordered based on the preferences of individuals at different stages. Younger individuals, more limited by income, appear willing to sacrifice amenities such as green spaces for city living. As these individuals move further along the life course they also move further and further away from the busy city centre. The interplay between preferences and constraints is apparent as income is seen to be one of the most limiting factors. As a result, the distribution of households across communities is affected by the general income of each household giving rise to working class neighbourhoods, middle class neighbourhoods, and student-communities among others. Thus communities become segregated based on the socio economic status of the households within them (PEACH, C., 1996).

Residential segregation occurs when specific groups live separately from one another despite existing in the same community (PEACH, C., 1996). Communities are mainly thought to be segregated due to differences in incomes, race and the structure of urban space (MASSEY, D. S. and Fischer, M. J., 2000). Segregation need not always be viewed as a negative feature of society as communities are often structured in such a way that stronger community groups and relationships develop. Not only is segregation linked to individual preferences but also due to the actions of private developers wishing to attract high income households. When segregation of this nature occurs, the problems associated with concentrations of poverty are not prevalent in these communities.

If communities are segregated based on the socio economic status of the residents within it, it is likely that deprived neighbourhoods will exist. This has been especially evident in the council sector where housing is reserved for those in dire need. Coupled with the problems of crime, low educational attainment and teenage pregnancy, a general lack of job opportunities can plague these neighbourhoods ((MASSEY, D S, 1996); (AGUILERA, A. and Ugalde, E., 2007)), and because of the high levels of unemployment, many households in such communities rely heavily on welfare support. Concentrations of deprivation as a result of these issues are deemed to be negative; they result in the social disadvantage and marginalisation of this vulnerable group (KEARNS, A. and Parkinson, M., 2001).

Due to the preferences of some households, segregation based on ethnicity may also occur. As mentioned earlier, minority groups may choose to live in areas where others like themselves exist, to buffer the effects of racial prejudice or the fear of it. Though this creates a comfortable community for minority groups, it can lead to the discomfort of long standing residents in the area causing these residents to move out of the community if they can. Through this process, formally known as white flight, communities become highly polarised based on ethnicity (JOHNSTON, R *et al.*, 2006). It is thought that the actions of estate agents also contribute to high levels of ethnic segregation across communities ((PHILLIPS, D, 1981); (PHILLIPS, D, 1988)).

Inevitably society is segregated for one reason or another. Though segregation is not always to be viewed in a negative context, when segregation leads to concentration of deprivation and marginalisation of households within a particular community, there is a cause for much concern.

2.3 Notes on the EASEL district and the wider Leeds Community

A comparison can be made between the residential mobility trends recorded in the literature and the trends observed in the EASEL district. The EASEL Housing Needs and Aspirations Study 2007 provides some basis for this comparison to be made (LEEDS CITY COUNCIL, 2007b). This study is notably based on a very small sample population of 166 households out of a possible 35000 EASEL area households. This is the only study available to date however, therefore the actual statistics will not be used here but rather the general trends observed will be highlighted.

In general, those living in council housing in the EASEL area express an interest in moving to other council housing in the surrounding district. Favouring family homes of at least 3 bedrooms, there is a general feeling that residents are satisfied with the access to goods and services though the desire to move may be fuelled by problems in the current neighbourhood, for example, crime. Other residents express an interest to move to be closer to family in neighbouring districts. There is not much consideration for access to transport routes and jobs as the majority of the residents interviewed were unemployed. Access to schools was also noted though this was not a general concern for the majority of residents. There is also not much interest in home ownership though the survey suggests that this is due to a lack of understanding of how the intermediate tenure market works; Local Authority housing schemes which allow tenants to rent with the goal of ownership are referred to as the intermediate market as they transition tenants from the public renting market to ownership (GJESSING, M., 2010).

Using the Leeds Housing Market Assessment 2006 more can be gleaned about trends in the general Leeds district outside of council housing (LEEDS CITY COUNCIL, 2007c).

This survey recorded a higher response rate than the EASEL Needs and Aspirations Study 2007. 3,543 residents responded and this instils more confidents in the results.

According to the Leeds Housing Market Assessment 2006, respondents noted that the choice of a new area takes into consideration the close proximity to work, schools, family, shops, leisure facilities, healthcare facilities and religious and cultural facilities. Households also move to a specific area because suitable housing is available; good quality housing. Areas where low levels of crime and antisocial behaviour are reported are also prime considerations. For Black and Minority Ethnic groups (BME), an area where there is no fear of racial and other harassment in the area is also very important. The consideration of house prices is very important, for households can only move if they can afford the new home.

In general, established households move to outer suburban areas and are likely to move to ownership because of their higher earning capacities. These types of households also move to larger homes. New forming households tend to rent in the private market and are likely to live in more central, urban areas. There is a general interest in the intermediate tenure market and demand for council housing remains.

Overall the trends established in the literature are comparable to those observed in the EASEL district and the wider Leeds area. In other words, there are similarities in the way that households in the public market and households in the private market make the residential mobility decision. For example, all households move to areas they can afford and take into consideration access to work, goods and services as well as neighbourhood quality. All households consider areas to which they are familiar among other factors. Thus, though clauses will be included to amplify the behaviour of social housing tenants, there will be no specific behaviours implemented in the quantitative analysis targeted at this group only. For example, the intermediate housing sector is not simulated.

The EASEL district is a large area with a high concentration of social housing. Amidst residential mobility trends and household preferences, the district is affected by various social and welfare problems which appear to be exacerbated with time. In light of this, there appears to be a drive by government to reduce the number of social housing provisions while allowing for an increased number of private housing provisions. Social housing tenants are encouraged to join the intermediate housing market which, over time, can lead to home ownership. For this reason, it is important that the policy background affecting areas such as these is clearly understood and the drivers influencing the change from social housing to private housing are discussed.

2.4 Housing Policy

Housing policy is used to regulate activity in the housing market and has evolved over the years in an effort to do so. Fundamentally affected by the supply of, and demand for, houses, the housing market is driven by competition between suppliers to satisfy housing consumption needs and the people who purchase houses on this market. For private developers, it is a market where profit making is the ultimate goal. Households will buy what they can afford, sacrificing preferred house features, such as number of rooms, in some cases. For those unable to compete in this environment, it is a market from which they are automatically excluded. Faced with few choices, such individuals are prime candidates to be forced into homelessness. It was for the benefit of this group in the market that social housing was established. It is a low-cost alternative for the working class. Today, though social housing continues to exist, the problem of poor housing and poverty persist in this sector, defeating the very purpose for which it was created – to provide decent and affordable accommodation for the working class. To an extent, the issues in working class areas like that studied in this thesis are tied up with the history of social housing in British cities, and to understand these issues the manner in which successive governments have tried to cope with them through social housing policies must be understood. This is not to say that these problems do not exist in private sector markets; however, the strong control that the government has on the social housing sector means that its development and management is a major branch of housing and regeneration policy. In particular, the release of social housing to the private sector is a significant development in the last thirty years.

The social housing policies of local government have tended, historically, to focus on build and fix methodologies, but these have done little to solve problems of disadvantage. As a result of this, in recent times, there has been a drive in urban regeneration policies (MEGBOLUGBE, I. F. and Linneman, P. D., 1993) to make the private housing market more accessible to all. Housing policy now supports a new wave of programmes aimed at enabling the poorest of households to compete in the private housing market. Mixed tenure developments are being proposed where communities are diversified by socioeconomic class. The intermediate housing sector has been created, affording low-income households the opportunity to begin their housing careers on the rental market, leading to a state of ownership in this 'rent to own' programme, reducing government's responsibility of providing homes. However, much can be learnt from the past; earlier policy attempts directed at improving poor conditions in the housing sector suggests that many of the problems of disadvantage were created by the policies themselves. A major aspect of this thesis is to examine whether this release of social housing into the private market is likely to introduce genuine diversification. More immediately, however, this raises certain key questions: How have past housing policies altered the dynamics in the housing market? Can urban regeneration-focussed policies really help to cure the ills of disadvantaged communities?

2.4.1 Earlier Attempts to Regenerate Disadvantaged Communities

Regeneration is defined as the renewal of lost physical and social vitality in a community (BRAMLEY, G *et al.*, 2004). Over the years, traditional housing policies have largely focussed on improving the quality of life in and around communities and may be described as regeneration schemes in their own right. Such improvements ranged from the creation of additional green spaces, renovation and/or the provision of dwelling houses. Thus by improving the physical layout of a community, new businesses could be attracted thereby creating additional avenues for employment. As a part of these regeneration-type initiatives, the social housing sector was created for those not able to compete in the private housing market. Today, modern variations of council housing have arisen in the form of Housing Associations and other charitable organisations. In general, these philanthropic groups are called Registered Social Landlords (RSLs) and fall under the umbrella of social housing – any subsidised housing provided by or on behalf of the government.

Early legislation such as the Lodging Houses Act (1851), Labour Classes Dwelling Houses Act (1866) and the Artisans' and Labourers' Dwelling Improvement Act (1875) gave Local Authorities the power to purchase and clear areas of unfit dwellings, condemning nuisance dwellings deemed to be unsuitable for human habitation (HOLMANS, A.E., 1987). Poor ventilation and water supply, lack of proper sewerage services and structurally unstable buildings contributed to these bad housing conditions (MULLINS, D. *et al.*, 2006). This legislation regulated the market so as to ensure that the prime suppliers of housing provided better homes for the working class and in so doing helped to improve social conditions at the time. This regeneration-focused legislation helped to improve the living conditions of those less fortunate and unable to improve living conditions on their own. Despite attempts to improve living conditions the poorest of the poor continued to suffer. Thus, the Housing of the Working Classes Act charged Local Authorities with the responsibility of creating and maintaining new and existing homes for which reasonable rents were to be charged (MERRETT, S., 1979). Though government was then charged with this responsibility, public provision of housing was seen as a last resort; satisfying the needs of only those in dire need. In a way the public market was its own type of residualised sector, as those occupying this housing were known to be very poor.

For a while the government aimed to improve the conditions of existing housing and built more housing where needed. The effects of the First World War changed this, however. With the introduction of the Housing and Town Planning Act (1919), government was forced to temporarily remove the responsibility of house building from the private sector until the market recovered in such a way that house prices were more affordable to the average worker (MALPASS, P, 2000). These new homes were not limited to council tenants and saw the reintroduction of many family houses. Much emphasis was placed on green spaces and neighbourhood amenities such as schools and shops (MALPASS, P, 2000). The Housing Act (1930) and the Town and Country Planning Act (1944) continued the programme of slum clearance started before the war, allowing Local Authorities to acquire land at less than market values for the purpose of development as well as demolishing privately owned houses in areas deemed to be slums. It was government's aim to reduce the problems of overcrowding, poor ventilation, sewerage and inadequate housing supply. With the building of three to five storey flats, rents were reduced to ensure that the poorest in society could afford homes. Importantly, these acts also saw the government taking more control over local building plans and where residences could be built, with a more rigorous set of quality demands required from private housing developers.

It was after World War II that the need for housing was further exacerbated. Fuelled by the widespread destruction, housing construction was again the main thrust of housing policy between the years 1945 and 1953 (MULLINS, D. *et al*., 2006). Again the Housing Act (amended in 1949) gave Local Authorities the responsibility of providing housing. This housing was available to anyone who could afford it and not just the working class (MERRETT, S., 1979). As a result, the social housing sector grew significantly. Traditional detached, semi-detached and terrace houses were the preferred housing option for families and the polarity among housing tenures became less pronounced as communities, mixed by socioeconomic status, were formed within public housing.

These housing policies were created to bring improvements to the housing sector by removing slums, retrofitting existing homes and created new homes for those in need. To a large extent, the policies responded to the demands of the market whether those demands required improvements to poor quality housing due to lack of maintenance or new housing provision as a result of the destruction of the war. In the case of the war, the dynamics in the social housing sector changed as government's provision of housing was no longer limited to poor people, rather such provisions were made available to anyone displaced by the war. Again, though not labelled regeneration policies, housing policies such as these helped to bring general improvements in neighbourhood conditions for those in genuine need.

Housing policy continued to respond to the needs of the market. In time, traditional houses and back-to-back terraces were no longer the preferred building option as the need to house large amounts of people in areas with limited space encouraged the construction of high rise flats, and tax incentives for builders at this time favoured the building of this housing type over any other (JONES, P., 2005). However, high rises were unpopular as the prefabricated design proved unsuitable for the cold and damp British climate, the concrete and flat roofed structures were prone to condensation, and they were also poorly insulated ((POWER, A., 1993); (STEWART, J. and Rhoden, M., 2003)). Long corridors, poorly lit passages obscured from public view, and other design features increased public expenditure for policing, court costs, prisons and social services (STEWART, J. and Rhoden, M., 2003). The tragic collapse of Ronan Point in 1968 did not serve to reassure government that this housing design was sustainable (POWER, A., 1993). Thus, the problems of the housing sector continued despite policy attempts to provide resolutions; quick fix solutions sacrificed quality for quantity. Such problems affected the social housing sector more so than the private housing sector where the market did more to act against poor quality housing.

By the 1970's social housing, as created by government in previous interventions, was failing. The problems in the sector persisted despite the continued efforts by the government over the years. The solution implemented was the Right To Buy Act (1980) which encouraged home ownership and depleted the social housing stock owned and managed by government (FORD, J. and Burrows, R., 1999). Giving tenants the option to buy their homes at a reduced cost saw the loss of a significant number of council owned homes; wealthier families and those who qualified for high discounts, because of their long tenancies, took the opportunity to become home owners (BRAMLEY, G *et al.*, 2004). There appeared to be an underlying notion that this would be the beginning of the end of public housing as government aimed to privatise many of its housing assets (FORREST, R and Murie, A, 1988). As council tenants took advantage of the Right to Buy Act, two significant developments occurred: the more attractive council properties were sieved out of the public market through house sales and the better-off council tenants became home owners. Though the Right to Buy policy increased home ownership significantly it also created a more residualised social housing sector. This, in turn, drew forth questions as to the purpose of social housing. Whitehead (1993) suggests that the housing market had failed and needed to be fixed; the provision of social housing by government was inefficient and not able to meet demands. The government believed that by encouraging individuals to the private sector, the social issues could be fixed (MEGBOLUGBE, I. F. and Linneman, P. D., 1993).

While greater home ownership might be regarded as positive, increasing economic and social engagement, the Right to Buy Act left Local Authorities with a mixed batch of properties. Local Authorities now owned a larger proportion of flats and non-traditional residences because the prime housing stock – mostly family accommodation – was bought up (MULLINS, D. *et al*., 2006). Also, the tenants who remained were those heavily reliant on welfare support; the unemployed, long-term ill and the disabled. As a result, poverty, high incidences of crime, antisocial behaviour and drug abuse became the tolerated norm across social housing estates causing greater divides between those living under social care and those living in other housing tenures (POWER, A. & THE JOSEPH ROWNTREE FOUNDATION, 1999). With the adoption of a new Housing Benefits system under the Social Security and Housing Benefits Act (1982), testing the welfare of applicants ensured that only those in dire situations could be housed by government and the concentration of the most marginal in society became more apparent. Bramley *et al.* (2004) point out that with a limited amount of stock and a rationing policy adopted, persons with more acute problems assumed a higher profile than others, though both groups claimed to have a need for housing. These changes in housing policy continued to narrow the scope of those qualifying for social housing making the intensity of social problems across Local Authority estates more prevalent.

As housing policy evolved in the early to mid 20th century there was a continued focus on increasing the housing stock and providing decent houses all falling in line with the remit of urban regeneration. A lack of a programme of sustained maintenance saw many of these buildings in poor conditions and in some cases in ruins. Repeated rounds of programme intervention did little to prove that these area-based building projects were sufficient to reduce the prevalence of poverty or provide sustained housing. On one hand, housing policy positively filtered out those able to survive in the private sector market but these same policies failed to successfully improve the core problems of the disadvantaged and failed the test of sustainability (MCGREGOR, A. and MacLennan, D., 1992). Traditional approaches to housing policy were losing their focus as a more holistic approach was being promoted (MULLINS, D. *et al*., 2006). Government was of the belief that by tackling neighbourhood problems in a comprehensive way there was a chance that investment would mean lasting improvement, an aim at the core of the early 1990s social policy plan for regenerating urban communities (BRAMLEY, G *et al.*, 2004).

2.4.2 A New Approach to Urban Regeneration

The realisation that traditional approaches to housing policy did not provide lasting benefits to those dependent on it caused a shift from directly providing housing to those in need to enabling or helping people into mainstream housing. Over the years, regeneration focussed policies have evolved. Originating in 1995, the Single Regeneration Budget shifted the focus of housing policy away from area-based housing-led approaches. The New Deal for Communities (NDC) followed in 1999 with an emphasis on community ownership and involvement, joined-up thinking, public private partnerships and a long-term commitment to real change (BRAMLEY, G *et al.*, 2004). More recent in the regeneration policy agenda is the National Regeneration Policy 2001 and the Housing Market Renewal Pathfinder Initiative 2002. In the case of the latter initiative, nine pathfinder projects were created as a means of targeting deprived communities in parts of the North and the Midlands. These Pathfinder projects were the government's commitment to effecting sustainable regeneration and were somewhat pilot projects used to test and refine the results and relationships of the social partners. Most recent is the Housing and Regeneration Bill (2007/08), introduced to provide guidelines for the regeneration process.

As earlier mentioned, regeneration is not a new area of policy. Miller (1959) mentions three approaches to neighbourhood regeneration; redevelopment, rehabilitation and integration. Redevelopment consists of the removal of existing buildings and the re-use of cleared land for the implementation of new projects (MILLER, M. J., 1959). Rehabilitation is largely concerned with improving deteriorated buildings by way of refurbishments, while integration may be a combination of the two. Traditional housing policy focused on redevelopment and rehabilitation but this proved to be unsustainable. Newer approaches to regeneration in the 1990s emphasise durability and use the above approaches to revive the economy of deprived neighbourhoods (MCDONALD, S. *et al*., 2009). This collapse plays out as a stagnation of the local economy as the decline of a local economy is directly linked to market failure.

There is difficulty in defining the term sustainable but it connotes a state of permanence and or durability. By rehabilitating people, reviving physical infrastructure and injecting life into the neighbourhood economy, disconnected social groups can be reconnected with mainstream society more permanently therefore creating some measure of sustainability. By providing avenues for academic and vocational training, improving the appearance of neighbourhood surroundings, providing adequate services and jobs, the persistent need for welfare support can be reduced. As surmised by Carley and Kirk (1998), sustainable regeneration is a process which ensures that activities today leave future generations better off.

Overall, the success of regeneration relies on joined up policies (KLEINMAN, M. and Whitehead, C.M.E., 1999) and requires the cooperation of each government department with an impact on the key problems of deprivation in disadvantaged neighbourhoods. Taking into account policies for employment, education, health, housing and their physical environment, the plans are intended to affect government departments directly engaged in developing these services. But with a policy with such a broad all-inclusive goal, creating successful, joined-up partnerships is likely to be a challenge.

2.4.3 The Impact of Regeneration

While regeneration offers positive results that inject life into communities, there is a lack of consensus as to whether these positive effects are sufficient to eliminate the challenges that threaten the success of this national policy area (KLEINMAN, M, 2000). There is evidence of the success of physical and economic regeneration plans in Liverpool where the improvements in the physical surroundings and an injection of life into the economy saw the city winning the bid for the European capital of culture 2008. Developments such as the Museum of Liverpool, Pier Head Canal Link and Mersey Ferry Terminal are some of the improvements of note. The social impacts however are in question.

Leeds Local Government believes that with the aid of mixed tenures, regeneration efforts are likely to be sustainable (GJESSING, M., 2010). This is because in theory, different classes of people have the potential to attract a wider range of new businesses and new residents. Mixed communities are such that owners, social housing residents and private sector renters are meshed together into one community. It is a mechanism to enact social mixing at the local level (TUNSTALL, R., 2003). These types of cross tenure communities comprise of a range of people at different socio-economic levels, with different lifestyles, values and attitudes (BRIDGE, G., 2002).

Proponents for this form of tenure diversification argue that mixed communities can contribute to a smaller concentration of unemployed people by attracting economically active households to previously deprived neighbourhoods (BEEKMAN, T. *et al*., 2001). Kleinhans (2004) also notes that diversification of tenure brings improvements simply through the presence of less deprived households reducing the incidence of problems across council estates and in turn improving the liveability score. Others argue that the dispersal of the disadvantaged over larger areas can aid in reducing the social costs of individual local authorities and this thinning out of deprivation can lead to improvements in socio-economic indicators ((UITERMARK, J., 2003); (ATKINSON, R. and Kintrea, K., 2000)).

However, it is not sufficient to argue that, by diversifying tenures as a consequence of regeneration, the problem of deprivation is thinned out or that statistical indices are improved. Such side effects may be favourable but do not help to prove that these neighbourhoods, which may see a temporary reduction in crime and other nuisances, are sustainable. The heart of the problem still exists; deprivation is merely spread wider and is thinned out on individual estates. This reduces the prevalence of nuisance behaviour when estates are examined individually but collectively the problem remains the same. Forrest and Kearns (1999) note that regeneration projects involving tenure diversification on predominantly council estates have the potential to exacerbate physical and social difference and increase social tensions. This may be a consequence of the fact that despite exposure, residents may not share core values and lifestyle habits (BEEKMAN, T. *et al*., 2001).

Bramley *et al.* (2004) view such exposure differently. They claim that this can improve the chances of developing social networks by bringing better-off, better-connected people into once deprived areas and can help to reconnect excluded individuals. To accelerate this process, locations where social networking takes place between owners and tenants need to be readily available. Such places include schools, local shops and/or pubs (BRAMLEY, G. and Morgan, J., 2003). Despite theorised results, research by Jupp (1999) concluded that in practice there is limited contact between owners and renters living in mixed tenure estates. These conflicting views are evidence to the fact that the true impact of mixed tenure communities, created for the purpose of regeneration, is still uncertain.

Not only is social networking hoped to encourage interaction between owners and tenants but it is hoped that owners become good role models for tenants. The role model function is defined in terms of people's attitudes and behaviour toward their home, the living environment and general aspirations, as influenced by someone held in high esteem (TUNSTALL, R., 2003). In his study 'The Truly Disadvantaged', Wilson (1987) provided evidence of social exclusion in segregated urban neighbourhoods as a result of a lack of role models provided by successful middle-class working families. Without having working examples of good family dynamics and positive values, Wilson (1987) argues that disadvantaged groups could not be improved. This theory is somewhat weak as it ignores the fact that role models may also exist outside of the immediate community. Also, households within mixed tenure developments where different social classes live on different streets are unlikely to share resources and effect serious change (JUPP, B., 1999). This not only limits interaction but also makes the process difficult to monitor, as it may be condescending to ask tenants whether they view middle-class residents as their role models.

There is also the notion that property-led regeneration plans are a form of gentrification ultimately acting to drive low-income families out of areas. However, Cameron (1992) in following the outcomes of regeneration plans in Tyneside and Newcastle, notes that social housing tenants can benefit from regeneration schemes because there are housing provisions tailor-made for them as part of the scheme. Furthermore, the introduction of a middle class who can purchase homes in newly regenerated developments is not in itself a negative thing as this is likely to attract new businesses and provide job opportunities for the unemployed. Turok (1992) adds that housing-led regeneration can also provide jobs as a result of the construction-related activity and has the potential to attract investment in the form of new businesses as well as improve the general appearance of the neighbourhood. Jones and Watkins (1996) caution, however, that there is a need for long-term investment in such projects by all stakeholders if the regeneration of run-down areas is to be a success.

In the context of the EASEL Redevelopment Plan, mixed communities are communities mixed by socioeconomic status. Unlike Jupp's (1999) suggestion, plans for the EASEL area suggest that though houses will be of various tenures they will all follow the same pattern. The distribution of owners, private renters and social housing tenants is planned to be of a pepper pot style with social housing tenants scattered throughout the new developments (GJESSING, M., 2010). Therefore it should not be possible to clearly differentiate one type of tenure from the other.

Arguments to promote and limit the use of tenure diversification as a part of regeneration policy are diverse (ATKINSON, R. and Kintrea, K., 2000). Whether the policy of creating mixed communities to bring sustainable regeneration will yield the required results is questionable - there are not sufficient results on which to base an informed judgement. What is known however, illustrates that the theory overshadows the practicality of the results.

Examining the results of housing policy over the last century it seems apparent that government is now relinquishing the role of housing provider to the private market where the ultimate goal of the housing career is home ownership. Over the years, there has been a continuous sieving of households, with buildings moving from social housing to ownership or private rental to ownership. In line with the notion of the housing ladder, and bearing in mind the view that households view ownership as an investment, the private rental and social housing markets are seen as temporary stop gaps on the upwardly mobile route to home ownership.

With the introduction of the intermediate sector, again more sieving can be anticipated. Local councils are likely to choose tenants who are able to pay the rental fees and are sufficiently able to sustain such an agreement until ownership. This satisfies the better off council tenants, enabling them to compete in a market which excludes them because of low wages. But with the ongoing plans to demolish poor quality council owned housing without plans for replacement (GJESSING, M., 2010), new mixed tenure communities are not sufficient in mass to provide housing for those in the council sector. If the better off tenants are sieved out to join new mixed tenure developments, where will the remaining council tenants go? Remaining council estates will retain the poorer council tenants, council housing will retain its stigma and the problems of residualisation will continue. How then are regeneration schemes helping the worse off council tenants? It seems likely that the very people social housing was created for are the ones who benefit least from such interventions. If the housing market is largely facilitated by the private sector, the goal of achieving profits will continue to make this market unattainable for those reliant on social care. Though new affordable homes rules now mandate that new developments must meet the standard of providing at least 30% cheaper housing, there seems to be a need for subsidised housing in the market more extensively. This is in direct conflict with the goals of private sector companies.

2.5 Conclusion

Residential mobility and housing policy are interrelated processes in the housing market. Housing policy is used to regulate activity in the housing market and as such is able to limit residential mobility behaviour. Over the years, several housing policies were created to meet specific needs of the day. Many of these policies helped those reliant on the government for housing in the social housing sector. Social housing is not devoid of its problems, however. Issues such as under supply, inadequate maintenance and residualisation challenge the success and viability of this housing tenure from the inception of this housing form to the present day (**Table 2.2**).

Recent research has highlighted the role housing policy has played in creating some of these problems, namely residualisation as a result of the Right to Buy Act (FORD, J. and Burrows, R., 1999). With urban regeneration schemes taking shape and with the problems of residualisation and segregation affecting social housing estates, a change in the state of council housing is anticipated. Sustainable communities are expected to be the results of such schemes but with few case studies available to provide the needed scope to critically analyse the success of such projects over time, the impact of regeneration policies is not clearly understood.

As regeneration plans unfold in the EASEL district seeking to tackle serious problems of social disadvantage among council tenants, the potential for improvement is greatly anticipated. However, whether mixed communities are a stable outcome from this policy implementation in such a complex market is questionable; the market is driven not only by political ideologies but also by individual preferences. **Chapter 3** presents a discussion on computational modelling, it forms the basis and justification for introducing a model of housing choice as a powerful way of exploring the likely outcomes of regeneration projects in this complex system. Through modelling residential mobility behaviour can be replicated at the individual and the results of interacting individual level behaviour examined on the aggregate level. In this way, the effects of regeneration policy can be explored. Thus whether regeneration policy can increase diversity within communities can also be examined.

Period	**Problems arising**	**Legislation Introduced***	**Key Objectives**	**Resulting issues**
Late 19th Century	Inadequate housing supply House prices not within the reach of the working class	Lodging Houses Act 1851 Labour Classes Dwelling Houses Act 1866 Artisans' and Labourers' Dwelling Improvement Act 1875 Housing of the Working Classes Act 1890 Housing and Town Planning Act 1919	Eliminate poor quality housing and overcrowding	Slums still exist Interventions interrupted by the world wars
1900s to 1970s	Widespread destruction after the war Urgent need for housing	Housing Act 1930 Town and Country Planning Act 1944 Housing Act 1949 Building Materials and Town Development Act 1952	Increase housing quantity and improve housing quality	House quality was sacrificed for the quantity of houses built
1970s to 1980s		Housing Act 1980 (Right to Buy)	Privatise some council owned housing stock	Residualisation of housing stock Shortage of council housing Social exclusion
*1990s to 2007***	Residualisation of housing stock Shortage of council housing Social exclusion	New Deal for Communities 1999 Housing Green Paper 2000 Housing and Regeneration Bill 2007/2008	Create sustainable communities through the introduction of mixed tenure housing	Uncertain

Table 2.2 Summary of the Evolution of Housing Policy; Key Objectives and association problems during each period

* 'Legislation Introduced' represents significant legislation relevant to this study

** 2007 is used as the cut-off point for this study

Chapter 3
Modelling Residential Mobility

3.1 Introduction

Much has been discussed on residential mobility and the influence of urban regeneration policy. Though this theoretical view is critical, further insights can be gained if a replica of these processes is modelled computationally. Computer models are abstractions of reality; they are explicit representations of specific phenomena in the real world which allow for the analysis of conceptual ideas (DAVIS, F. D. *et al*., 1989). Through computer modelling, core dynamics of a specific phenomenon can be illuminated allowing for the examination of causal relationships (CROOKS, A *et al*., 2008).

Computer models can be used to optimise processes and/or simulate behaviour in the real world so as to focus policy arguments, highlight new questions, challenge accepted theory and unearth complex dynamics in seemingly simple systems (STERMAN, J. D., 1988). Optimisation models are largely prescriptive in nature; they are used to improve an existing process or find the best way to execute a process. For example, optimisation models may be used to provide solutions to '*the travelling salesman problem*'. Here, a model can be created to find the shortest path for a salesperson to travel when he/she visits several locations. On the other hand, simulation models are descriptive in nature; they are used to describe real world events by simulating real world systems at different time periods under changing conditions. This may be likened to what-if scenarios (HILL, R. R. Mcintyre, G. A. Narayanan, S., 2001). Models such as these use decision making rules to govern behaviours, the results of which are descriptions of modelled decisions. By using simulation models, modellers are able to systematically alter model parameters and compare results in a way not possible in reality.

Models have long been used to explore themes in the realm of ethnic segregation, land use, transport and retail planning as well as meteorology and flood risk management amongst others ((BANTON, M., 1994); (PARKER, D., 2006); (ZERGER, A. and Wealands, S., 2004)). Such models include those by Schelling (1969) who illustrated how slight preferences could result in the total segregation of a community. Yin (2009) built on Schelling's model and illustrated that when some housing policies were implemented segregation could be reduced if racial sensitivity was low. Each of these examples addresses spatial problems; they are deterministic, if with stochastic elements, and dynamic models. Conversely, models may be static, and/or mathematical in nature, the latter often used to investigate questions in the field of economics.

Amidst the usefulness of modelling, it is not without its challenges; the outcomes of computer models are said to be difficult to validate (CROOKS, A *et al*., 2008); their predictive power is often questioned (STERMAN, J. D., 1988); simulation modellers are faced with the challenge of accurately choosing and quantifying soft variables documented in the qualitative literature and explicitly defining behavioural rules derived from the same; also, defining model boundaries is noted to be a challenge in spatial modelling (STERMAN, J. D., 1988). Despite these challenges, it must be reiterated that computer models are abstractions of reality; they are used to explore specific aspects of the real world. Though faced with challenges, modelling provides a controlled environment where real world problems, conceptualised only in thought, can be examined in a way not possible in reality. When predictions are intended, however, model outcomes must be considered in the context of any assumptions made.

Simulation models can be used to recreate residential mobility behaviour and examine the influence of urban regeneration policies. To date, there exist some residential mobility models which have been created to address various themes within this topic area (**Section 3.3.2**). Recalling the original research question introduced in **Chapter 1**, '*What insights can computational modelling give about the success of the urban regeneration policy schemes enacted in the EASEL district?*', this chapter will be used to demonstrate that existing residential mobility/urban regeneration models are not able to adequately address this question. Various types of modelling techniques will be presented complemented by simple examples of existing documented models. Relevant models of residential mobility and/or urban regeneration policy are then reviewed in more detail. Finally, the modelling technique of choice is applied to the residential mobility theory discussed in the previous chapter, **Chapter 2**. It is the introduction to the rudiments of a new model of residential mobility, the CHAIRS simulation.

3.2 A Description of Contrasting Modelling Techniques

There are at least four contrasting spatial modelling techniques that may be used to explore the residential mobility problem posed in the research question. These techniques include spatial interaction modelling, microsimulation modelling, CA modelling and ABM. Each of these modelling techniques will be discussed in this section complete with an example of an existing model or real world system. ABM is identified as the modelling technique of choice because of its ability to replicate dynamic behaviour at the individual level. **Table 3.1** below is used to highlight the differences between each modelling technique.

Characteristics	**Spatial Interaction**	**Microsimulation**	**Cellular Automaton**	**Agent-based Modelling**
Main Purpose	Explanation, Projection	Projection	Explanation	Explanation and Prediction
Building blocks	Aggregate	Individual	Individual	Individual
Applications	Store location; retail planning	Policy implications, population prediction	Urban growth; physical analysis	Theory formulation, verification
Investigation focus	Spatial flows	Aggregate trends	Aggregate trends	Emergent behaviour
Communication between agents	No	No	Yes; however, no movement	Yes

Table 3.1 Comparison between four modelling techniques adapted from *Table 1* in Mahdavi *et al.* (2007; 367)

What is important to note is that spatial interaction modelling focuses on aggregate flows between a source and destination. This type of modelling does not drill down to the individual level. Though microsimulation modelling drills down to the individual level, there is no interaction between individuals therefore meaning that the behaviour of one individual does not rely on the behaviour of others. Like microsimulation, CA modelling drills down to the individual level and there is communication between agents, however, with CA modelling there is no movement of individuals from one location to another. ABM is an individual level modelling technique where communication between agents can be facilitated. Also, individuals can move from one location to another.

3.2.1 Spatial Interaction Modelling

Spatial interaction models are used to examine the movement or flow of real world objects between sources and destinations (FOTHERINGHAM, A. and O'Kelly, M., 1989). These objects may include people, goods and services, among others. Interaction flows are normally estimated based on indicators such as distance and demand (BAILEY, T. and Gatrell, A., 1995). The number of interactions is a function of the attractiveness of a destination from the origin and the impedance between the two locations; it is a tradeoff between attractiveness and accessibility (HAYNES, K. and Fortheringham, S., 1984). Spatial interaction models can be used to access a wide array of research questions estimating in/out migration between small areas – local communities, and larger domains – country to country; retail planning to identify the most profitable location for a new store and/or access the current performance of a particular brand of stores; and to evaluate the demands for other services such as transport, assisting in transport planning. In this way, migration totals can be assessed, profits and losses can be estimated and movement counts analysed.

Residential flows are an important element of many spatial interaction models. A policy relevant application involving residential flows is presented by Clarke *et al.* (2002) in which a model of spatial interaction is used to access the impact of opening new food retailing stores on food deserts. Here food deserts are described as areas where there is poor access to food outlets. The spatial interaction model is used to quantify the level of interaction between residential zones and grocery stores. In this way, the potential performance of grocery stores can be reported. Interactions are calculated using a mathematical function taking into account many factors such as expenditure by household type, attractiveness to grocery retail destinations, distance decay parameters by household type, distance between residential zones and grocery stores, average weekly expenditure by household type, number of households in each residential zone and average revenue per square foot of grocer floor space. Results from this model were able to show how the closure of Netto and Co-op stores in Seacroft, Leeds could affect the local area; though these closures only decreased store access of households by 13%, the results of the model suggests that the closure of these two low cost grocery stores would likely affect lower social classes; those with a high probability of being less mobile (CLARKE, G. *et al*., 2002).

Likewise Hincks and Wong (2010) present a spatial interaction model of housing and labour market interaction. Using a case study of North West England commuting flows between housing market areas were analysed. Housing market areas are noted to be areas in which households may search for a new home such that changing jobs is not necessitated (HINCKS, S. and Wong, C., 2010). Using a case study of North West England, housing market areas that were coincident with labour market areas tended to be the one with the highest commuting flows. Much like the previous example of the work by Clarke *et al*. (2002), residential flows are replicated in this example all at the aggregate level. Thus the modelling of individual level dynamics is not facilitated when spatial interaction models are constructed.

3.2.2 Microsimulation Modelling

Microsimulation is a stochastic, statistical procedure for estimating the characteristics of individuals based on knowledge of the aggregate characteristics of the population to which the individuals belong (JOHNSTON, R., 2000). Clarke (1996) describes microsimulation as a deductive technique used to model society by reproducing demographic, social and/or economic characteristics of human behaviour. Microsimulation modelling is therefore used to examine aggregate trends over time. Unlike spatial interaction models, microsimulation operates at the individual or micro-unit level; individuals include entities such as people, households and organisations and are used to project future events such as population growth and the impact of public policy. Dynamics at the individual level are of average complexity, usually based on the statistical rolling-forwards in time of populations, and there is no interaction between individuals (MAHDAVI, B. *et al*., 2007). The fact that there is no interaction between individuals is significant when the comparison is made to ABMs as the latter technique allows for interaction at the individual level. Therefore, though microsimulation models are able to replicate individual level dynamics, the added real world functionality of simulation interaction between these individuals cannot be facilitated. Two examples of microsimulation models are presented, the first applied to the housing market and the second used to illustrate how government policy can be analysed.

Wood *et al.* (2006) use the microsimulation technique to create a model of the Australian housing market focusing mainly on tenure choice, that is, ownership or rental. The model considered house prices, household wealth and borrowing constraints related to each household when the decision to find a new house was to be made. The model was also used to assess the effects of government interventions when these variables were altered. Thus supply and demand functions of the housing market are modelled using the microsimulation tool. Individual level behaviour is largely based on the willingness of a household to pay a specified price for a new home. Nothing that households have a fundamental preference for ownership, constraining factors such as wealth and access to credit facilities are all part of the behavioural functions included. Though the model results showed that there was a high demand for home ownership, households were largely limited by the large down-payments required by lending agencies. The writers concluded that with the help of government grants to assist with the required down-payment, some households could be guided to home ownership.

In a similar way, Ballas and Clarke (2001) explored the impacts of major national policy related to government benefits such as unemployment and housing benefits as well as jobseekers' allowance on households at the small area level. The model, developed for the Leeds urban system, explored the spatial impacts of changes in taxation policy and child benefit policy on households (BALLAS, D and Clarke, G, 2001). Where the child benefit policy is concerned, the model, SimLeeds, illustrated that areas with the highest amounts of child benefits paid are the most affected when this policy is altered. Though this stands to reason in theory, the model is used to explore the extent to which raising and/or lowering child benefits could impact households, an exercise not automatically possible in reality. This is possible by assigning behaviours and attributes to households at the small area level and producing outcomes as a result of policy implementation in the form of simulated scenarios. The techniques of iterative proportional fitting and simulated annealing reweighting in collaboration with conditional probabilities derived from the household Sample of Anonymised Records (SAR) and the small area statistics (SAS) were the two methodologies used to create the base population and estimate behaviour throughout the model. These techniques are defined in greater detail in the article by Ballas *et al.* (1999). The household and individual level attributes used in the system included age, sex, marital status, housing tenure, and employment status.

3.2.3 Cellular Automaton Modelling

CA modelling is a discrete, deterministic modelling technique (BENJAMIN, S. C. *et al*., 1996). CA models are made up of cells in a tessellated grid space which may be likened to squares on a checkerboard. The state of each cell is governed by a set of simple rules applied iteratively over a period of time and is based on the state of current cells and neighbouring cells (AL-RABADI, A, 2011). Each cell in the CA grid space may be used to represent entities such as neighbourhoods and households, in this way, spatial simulation can be carried out. Thus CA models can be used to analyse residential mobility and other urban phenomena. Though CA models operate at the individual level and can be linked to a geographic information system to represent real life phenomenon, they tend to be abstract as with other modelling techniques (MAHDAVI, B. *et al*., 2007). Also, there is no movement of individual entities in this type of modelling.

The model created by Caruso *et al.* (2007) is used to illustrate how CA can be applied to the housing market. Using households and farmers as the major entities of interest, each cell in the CA grid space corresponds to a residential plot and is occupied by either a household or a farmer. Farmers commutes to the central business district (CBD) in order to sell produce while the head of each household commutes to the CBD to buy goods as well as to work. In this way, commuting costs and other expenses are the factors that constrain these households as they are limited in terms of their ability to pay rent based on the amount of money available to finance these expenses. As a result of this rent adaptations across the landscape are effected to match what individual households can afford at a given time. In this way, the characteristics of neighbourhoods change over time. The underlying economic relationship in this model are driven by an economic model. Economic models are discussed in a subsequent section, **Section 3.2.5**.

CA models can also be used in simulating traffic systems. For example, Benjamin *et al.* (1996) analysed traffic flows along a highway. Defined on a simple *N* by *N* lattice grid, this CA model illustrated how the presence of a road junction could improve traffic flow. With each grid cell having a state of occupied or vacant, all movement was governed by a set of basic rules; each vacant cell assumed the state of the cell on its immediate left, and each occupied cell assumed the state of the cell on its immediate right (BENJAMIN, S. C. *et al*., 1996). In addition, the acceleration, disorder and slow-to-start rules governed the speed of each vehicle based on probabilistic outcomes. Here the disorder rule was used to account for the sometimes irrational behaviour of humans. Results of the model illustrated that the when a junction is added to the roadway it is most effective when there is disorder on the highway. Also, the performance of the junction could be improved when the speed of vehicles on the highway is reduced.

Though this model does not make use of actual data in real life geographies, urban planners may opt to use a tool such as this in accessing the effectiveness of junctions. In this case, a model such as this lends support to the use of junctions on long highways. CA models have also been used for land-use modelling concerned with urban spread and rural-urban transitions (CAGLIONI, M. *et al*., 2006). Each of these applications include features regarding different aspects of the housing market rather than the issue of housing reallocation. Again it is important to reiterate that though CA modelling is an individual level modelling tool, there is no movement of individuals within the model. This observation can be contrasted with ABMs where there is movement within such models.

3.2.4 Agent-based Modelling (ABM)

ABM is an artificial intelligence technique which utilises and implements concepts often applied to humans (1997). Such models encapsulate the themes of multi-agent systems, which use a set of individual, intelligent, interacting, decision-making agents to recreate real world behaviour (RUSSELL, S. and Norvig, P., 2003). Multi-agent systems are often described as complex systems; systems which are made up of autonomous individuals interacting in a nonlinear way thus giving rise to emergent behaviour (AXTELL, R. and Epstein, J. M., 1994). Here, emergent behaviour is defined as unexpected patterns of behaviour resulting from the interaction of simple rule-based individuals (BONABEAU, E., 2002). Agent-based models (ABMs) are dynamic, deterministic systems and can be used to explain a specific phenomenon (AXELROD, R, 2006).

Like microsimulation modelling, ABMs allow for the manipulation of entities at the individual level. These entities are often called agents. As a result of this, ABMs can be used to represent human behaviour in a way similar to reality. At the individual level, agents are driven by simple behavioural rules which govern how they interact with each other and with the environment around them; agent states can change due to their interaction with other agents; agent states can also change due to interaction with the environment in which they exist. When agents are observed collectively, the resultant behaviour is more complex. This observation is at the heart of ABM, that is, simple individual level behaviour can produce complex behaviour when aggregated; emergent behaviour (BONABEAU, E., 2002).

One example of a complex system that can be represented using ABM is that of an ant colony carrying out its daily duties. Noting that each individual ant (agent) must ensure that the task to which it was originally engaged is not oversubscribed, the emergent behaviour created is a balancing off of ants (agents) in each group; after some time each group will contain on average the same number of ants (agents). Collectively, the ants appear to work as a single organism, yet there is no leader guiding each agent towards its role (RESNICK, M., 1997). Instead, each ant agent reacts to the behaviour of its immediate neighbour who in turn reacts to it. This results in a balanced community. It is this emergent behaviour that helps to explain the behaviour of the ant colony. Though behaviour is simple at the individual level the collective behaviour caused by the continual interaction between simple rule-based agents creates complex resultant behaviour (JENNINGS, N. R., 2000).

Agents are generally programmed using an object oriented programming language such as Java where a special-purpose simulation library or modelling environment can be created. Individual-level behaviour is driven by collections of condition-action rules which agents use to perceive and react to their situation in pursuit of an ultimate goal (MACAL, C. M. and North, M. J., 2010). These rules provide a mechanism for agents to communicate.

The agent-based methodology can be used to recreate residential mobility behaviour. A model can be created which uses households as individual agents. Households can be assigned attributes such as age, housing tenure, social class. The neighbourhood environment may include letting rates, job opportunities, neighbourhood quality and mechanisms can be included to cause neighbourhood change over time. The financial environment may include mortgage conditions and lending facilities as well as interest rates. Again, a mechanism would exist to alter these conditions over time. In a similar way, aging models can be included and linked to mortality, fertility; this by extension influences changes in the family life cycle. Households interpret these environmental conditions based on various behavioural rules. Thus by using this constructed environment, each household agent can make the residential mobility decision by observing changes in the surrounding environments as well as changes effected by the relocation decisions of other households. What results from this repetitive, competitive interaction between simple rule-based agents is a reshuffling of households from house to house. Aggregate behaviour observed over time is likely to reveal trends in the distribution of household most notably, patterns of segregation. Much like microsimulation models, ABMs can be constructed in a way that policy scenarios can be implemented, in this case answering policy focused questions related to residential mobility and urban regeneration.

If agent-based models attempt to mirror real world systems then the way in which agents make decisions is very important. In the previous example, agents may make decisions in various ways; for example, decisions may be made based on a weighted combination conditions. However, more formal decision making mechanisms exist to govern the way in which agents reason. According to Panzarasa *et al.* (2001), the Beliefs Desires Intentions (BDI) model is by far the most popular and influential model of cognition associated with the agent-based approach. Bratman *et al.* (1988) describe this framework as one in which each agent weighs competing alternatives in a process of means-end reasoning. The system is such that beliefs form the informative component of the system state; they are the agent's knowledge of the world around them (RAO, A. and Georgeff, M., 1995). Desires are the objectives to be accomplished; goals which can be used to represent the motivational state of the agent. While goals which have been committed to are the intentions (MÓRA, M. C. *et al.*, 1999).

Other techniques such as Schmidt's Physical Conditions, Emotional State, Cognitive Capabilities and Social Status (PECS) behaviour framework are also used to simulate the way in which agents make decisions. Schmidt (2000) argues that the PECS framework is more representative of the way in which human beings function as it takes into account the physical condition, emotional state, cognitive capabilities and social status of each individual. Malleson *et al.* (2008) presents a model of burglary where the individual behaviour of burglars is modelled using the PECS frameworks and integrated into an ABM. In the case of the BDI model, much emphasis is placed on the cognitive capabilities of individual agents while the physical condition, emotional state and social status are often represented at a higher level of abstraction than issues of cognition.

Whether one of these systems is used to direct decision making or a bespoke model of decision making is used, it is very important that the behavioural rules are calibrated to mirror similar behaviours in reality, that is, rule elements that are weakly representative of the real world are adjusted so the model gives the best possible reproduction of reality. Since agents are nonlinear and often quasi-stochastic; their nature is not always predictable; a major characteristic of human interaction. Calibrations and the associated validation (comparison of the model results with reality) are often difficult when examining the results of ABMs (CROOKS, A *et al*., 2008). Nevertheless, calibration ensures that any emergent behaviour predicted by the model occurs despite stochastic and non-linear elements in the model. The difficulty chiefly lies in finding sufficient data to use in the adjustment and validation processes. Calibration and validation are discussed in detail in **Chapter 5**.

3.2.5 Why Agent-based Modelling?

Table 3.1 can be used to compare and contrast the main features of each of the mentioned modelling techniques. The models can be analysed in terms of the purpose of the model, building blocks, usefulness in terms of applications, investigation focus and communication between agents. Microsimulation models are generally used for projections. Using this modelling technique, demographic details of a population can be projected, namely population growth in a particular geographical area. Spatial interaction, CA and ABM are generally used to explain a specific phenomenon in a system. These modelling techniques may be used to answer questions related to urban sprawl, land-use change, residential segregation and other physical/social problems. Also, whereas spatial interaction models operate at the aggregate level, describing spatial flows of various entities, the other modelling techniques operate at the individual level though with varied dynamics. For example, though operating with micro-units at the individual level, micro-units do not communicate with each other using the microsimulation technique. Conversely, at the individual level there is interaction between micro-units when the CA and ABM techniques are used.

Characteristics	Spatial Interaction	Microsimulation	Cellular Automaton	Agent-based Modelling
Main Purpose	Explanation, Projection	Projection	Explanation	Explanation and Prediction
Building blocks	Aggregate	Individual	Individual	Individual
Applications	Store location; retail planning	Policy implications, population prediction	Urban growth; physical analysis	Theory formulation, verification
Investigation focus	Spatial flows	Aggregate trends	Aggregate trends	Emergent behaviour
Communication between agents	No	No	Yes; however, no movement	Yes

Table 3.1 repeated Comparison between four modelling techniques adapted from *Table 1* in Mahdavi *et al.* (2007; 367)

Most of these modelling techniques can be distinguished from each other based on the review in **Table 3.1** as repeated above. However, there are clear similarities between the CA and ABM techniques. The difference between these modelling techniques lies in the way in which agents interact in both models. That is, in ABMs there is some level of asynchrony among agents; agents do not necessarily perform actions simultaneously (MACAL, C. M. and North, M. J., 2010). Governed by specific rules, agents make decisions based on the feedback gauged from other agents and the environment in which they exist. Each agent senses the environment in which it exist and reacts to it based on the agent's goal (FRANKLIN, S. and Graesser, A., 1996). Agents are also described as being autonomous and flexible (JENNINGS, N. R. *et al*., 1998), in that they can act on their own volition, learn from past behaviour and alter their decisions based on present circumstances. In general, agents can be self-learning, independent, goal-oriented and exhibit a measure of rationality in achieving each goal ((CASTELFRANCHI, C., 1998); (RUSSELL, S. and Norvig, P., 2003); (BONABEAU, E., 2002)). This is unlike CA models where agents make decisions simultaneously (HEATH, B. L. and Hill, R., 2010). Furthermore, in CA modelling there is no movement of individual entities, an outcome possible in ABM. Also, unlike CA models which are limited to a grid-based set up, ABM models are not limited in this way; they can be applied to real geographies using both raster and vector details (AXTELL, R. and Epstein, J. M., 1994).

Amidst the benefits of ABM, it may be argued that statistical methods employed by economists could be considered as a viable methodology for building a model of residential mobility. Such models can be used to identify solutions for economic problems. In the realm of residential mobility, economic models have been used in the context of analysing the impact of changing house and land prices when compared to household income and/or socioeconomic status. Some of these models include the work of Goodman (1976) where the conclusion is drawn that when housing consumption does not match housing utility then a local move can be effected. In this case, utility is a function of house prices and disposable income among others. Like ABMs, economic models provide a mechanism for qualitative theory to be analysed empirically, this can be observed in the work of Alonso (1964), Hanushek and Quigley (1978), Quigley and Weinberg (1977), Lee and Trost (1978), Goodman (1995), Ihlanfeldt (1981), Senior *et al.* (2006) and Furtado (2008) among others.

Though economic models are able to replicate existing statistical relationships, unlike ABMs, economic models work at the aggregate level and this is an obvious disadvantage if individual level dynamics are to be simulated. Though a hybrid model of some kind could be created to merge the benefits of economic modelling with ABM, there is still merit in replicating individual level dynamics using ABM only, most notably the fact that such a model does not yet exist. Noting this, a hybrid model could be a consideration for future work, incorporating the expertise of an economist.

Overall, ABMs allow for real world dynamics to be included in the execution of the model in a way not possible with any of the other modelling techniques. This is a highlight of ABMs as these dynamics are not possible in spatial interaction or microsimulation models and are limited in CAs. For this reason, ABM is the modelling technique of choice for use when creating an appropriate residential mobility model to answer the original research question posed.

3.3 Comparing and Contrasting Existing Residential Mobility ABMs

3.3.1 Thomas Schelling's Model

The work of Thomas Schelling is noted to be one of the first agent-based models of its kind to replicate discriminatory residential mobility behaviour in the form of a model (MAHDAVI, B. *et al.*, 2007). As earlier mentioned, Schelling ((1969), (1971)) examined the role of preferences in an artificial community and illustrates how simple behaviour at the individual level can create significant collective results not directly intended by the individual (SCHELLING, T. C., 1969).

Schelling used 70 individuals randomly distributed across a line. The model used arbitrary objects; 35 crosses, 35 circles, which represented individuals in a neighbourhood. Individuals were noted to be satisfied if they lived around other individuals of the same type. Unsatisfied individuals moved to a randomly selected new location. Satisfaction was therefore based on the individual's preference for living among others like themselves. The results of this model showed that even when neighbours had slight preferences to live among others like themselves, total segregation was effected. In the real world, such preferences may be based on individual attributes such as ethnicity, social class and housing tenure.

Figure 3.1 below is an illustration of Schelling's model. Note that the plus sign '+' represents the crosses in the above description. The first line represents the distribution of individuals before the rules are applied and the dots indicate the individuals that are not satisfied. The second line shows the results when the rule is applied.

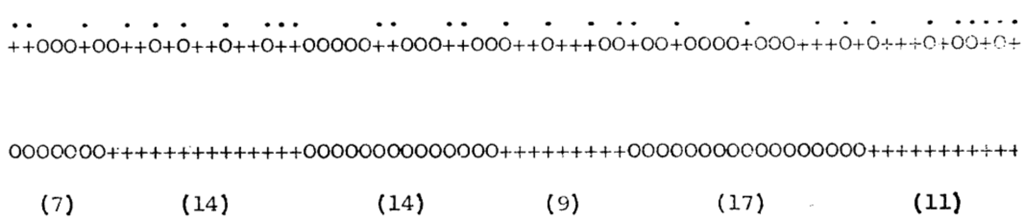


Figure 3.1 Schelling's 1-D work space as originally illustrated in Schelling (1969)

The results show that when the basic preference rule is applied, distinct clusters of like neighbours are generated as shown in line 2 of **Figure 3.1**. In a similar way, Schelling extended this work to apply to two-dimensional checkerboard space such that a more practical parallel could be made to the real world. Using this environment, segregation trends were still observed (SCHELLING, T. C., 1971). In general, Schelling's theory suggests that even when individuals were willing to be in the minority, segregation was a stable equilibrium while integration was not (O'SULLIVAN, D, 2009).

Schelling's work was the foundation of understanding residential mobility dynamics at the individual level, however, this model is limited. Schelling's model was not linked to actual observed research based on real populations. The model was not applied to a real geography and as a consequence there is uncertainty as to whether the segregation patterns realised by Schelling match those in reality. Also, Bruch and Mare (2006) note that residential mobility behaviour is generally derived from statistical analysis of Census data and soft variables described in the qualitative literature. Information such as this is useful, however, ABM allows for individual-level dynamics to be further examined.

By applying Schelling's model to an actual geography some of these reservations can be tested. Using a subset of 180 houses in the EASEL districts, households generated through microsimulation (**Section 4.2.2.2**), were randomly assigned to each house. In this illustration there are 163 households and 17 vacant houses. Detailed characteristics of the neighbourhood, houses and households are not of importance here, instead two types of households exist; red and yellow households, where the colour variation is symbolic of any one household attribute; race, income, social class etc. Using this real geography, households in the model are accessed. For each household, all households living within a 100m distance was analysed. If more than 33% of the neighbouring households were not of the same type, the household in question was noted to be unhappy with its present location. Note that 33% is the tolerance threshold first identified in Schelling's model (SCHELLING, T. C., 1969). Once households were noted to be unhappy they were moved to any randomly chosen vacant house. The figures below illustrate the results of this modelling process. **Figure 3.2** shows the model when randomly initialised with households while **Figure 3.3** shows the resultant population distribution once the preference rule is applied.

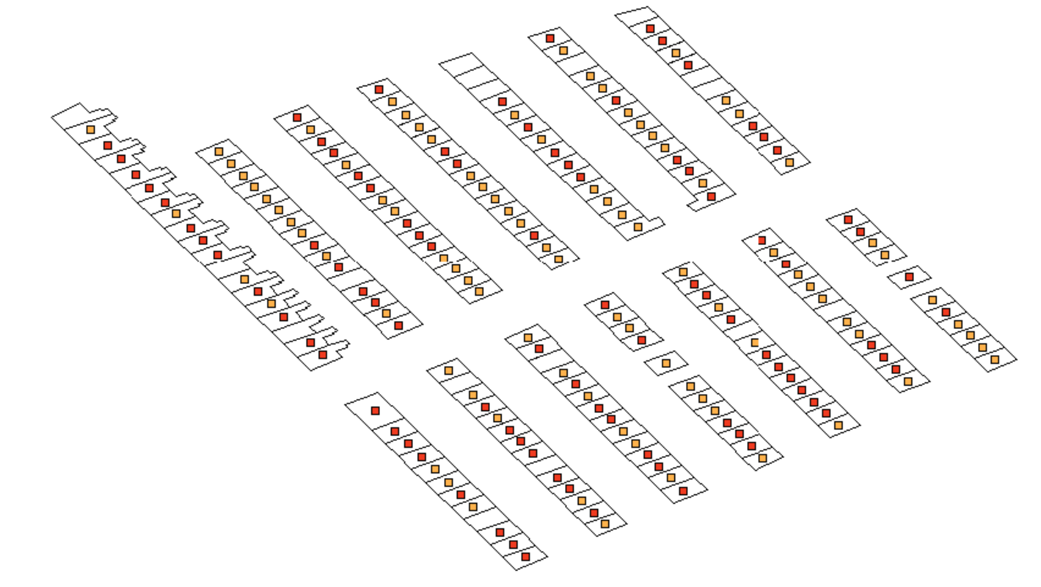


Figure 3.2 Initialised population randomly distributed

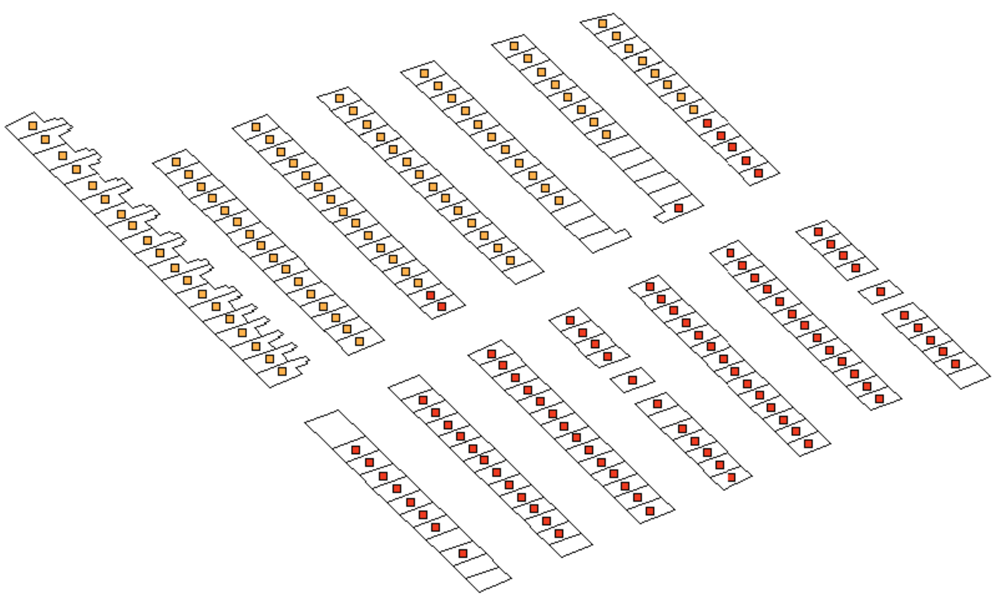


Figure 3.3 Population distribution when residential mobility rule is applied

Though individuals are not likely to move to a randomly selected new location when relocating, the results of this basic illustration show that when applied to an actual geography, Schelling's theory is still applicable; segregation is still realised. Similar observations are made in the work of Crooks (2010) who recreated Schelling's model using an actual geography. Nevertheless, Crooks (2010) noted that neighbourhood size and geographical features were important considerations when segregation patterns were analysed highlighting that the geometry of a specific area could act as a barrier to segregation. Such an observation illustrates the importance of using real geographies in modelling residential mobility behaviour.

Based on the qualitative literature cited in **Section 2.2**, the dynamic behaviour at the individual level is thought to be based on several factors and therefore a complicated number of decision-making processes are likely to be experienced by each household. This is one of the reasons that work by other researchers have used Schelling's basic model to build models of residential mobility but have used increased complicated dynamics.

Research by Laurie *et al.* (2003), Bruch and Mare (2006), Zhang (2004), Aguilera and Ugalde (2007), Benenson (2004), Fosset (2009), Pancs and Vriend (2007) and Yin (2009) are some of the more recent attempts at analysing and extending Schelling's earlier work. Each model aims to recreate residential mobility behaviour and has been created in different ways with varied foci. Much of this work is discussed in the section to follow as a means of establishing the extent to which the residential mobility models have been created using ABMs and assessing whether such models are sufficient to answer the research question posed.

3.3.2 More Recent Models of Residential Mobility

There are five residential mobility models presented in this section. These models are part of the body of more recent documented research in this subject area. Though other models could be presented (FOSSETT, M and Waren, W, 2005); (PANCS, R. and Vriend, N. J., 2007); (FOSSETT, M. and Senft, R., 2004), it should be noted that each of these models share one commonality; they only incorporate at most three mobility behaviours. This is significant in the context of the CHAIRS simulation which is presented in **Section 3.4**. In stark contrast to other residential mobility models, the CHAIRS simulation explores the use of seven mobility behaviours. A summary of the models presented in this section can be found in **Table 3.2**, it follows this discussion.

The work of Laurie *et al.* (2003) used the basic segregation model as proposed by Schelling ((1969), (1971)) in examining the role of vision in effecting segregation. Vision is the term used to describe the number of neighbours each agent assesses in determining whether he/she is satisfied with the current neighbourhood. It is called the *R*-neighbourhood. The aim of the model was to explain how variable values of *R* affect the nature of segregation. Two types of agents exist in the model; black and white. Initially, agents are randomly distributed across a two-dimensional lattice grid. Each agent first observes its *R*-neighbourhood, if the number of like agents is greater than or equal to some predetermined preference level then the agent is noted to be satisfied, else the agent is noted to be unsatisfied. This is, in effect, Schelling's original model. Unlike Schelling's model which simply randomly relocates unsatisfied agents, in Laurie *et al.*'s model, unsatisfied agents randomly select vacant cells and conduct the *R*-neighbourhood assessment once more. This process is repeated until the agent finds a suitable new location or the threshold for searching is exhausted. Overall, the evaluation process in the model is repeated until all agents are satisfied with their locations meaning that equilibrium is reached.

When the model is initialised with agents in random locations, the results suggest that when the *R*-neighbourhood was small and agents preferred to live around 50% of neighbours like themselves, segregation occurred in partially connected ghettos. As the value of *R* increased, total segregation occurred. However, for communities with lower preference percentages, as the *R*-neighbourhood increased, segregation decreased. These findings can be extended to policy, suggesting that as the preference for like neighbours decreases and vision is increased, stable, integrated neighbourhoods can be created. Thus public policies geared toward increasing vision could lead to more integration within communities as the preference percentage is low.

Bruch and Mare (2006) built a model of residential segregation where the main aim was to examine how residential mobility behaviour could likely produce segregation patterns in neighbourhoods of Los Angeles. The model used behavioural rules based on race and income. Bruch's model was initialised with renters only, households moved based on the calculation of transition probabilities derived from census data while house prices were updated based on a function of neighbourhood turnover rates.

In Bruch and Mare's model, 5% of the households and 50% of vacant houses were chosen randomly and used to simulate the relocation process in an actual geography. The system did not account for population growth in the form of natural increase or immigration and was therefore representative of a closed system. Also, the model was not validated against real data as a mean of testing the goodness of fit of the results. Overall, the model showed how residential mobility patterns based on income were likely to decrease segregation relative to residential mobility patterns based on race where increased segregation was observed.

Zhang (2004) used a checkerboard lattice grid to recreate residential mobility behaviour. This mathematical model was used to show that residential segregation persists even if every individual in society preferred to live in a half white, half black neighbourhood. The model suggests that residential segregation could persist in a society regardless of what individuals want purely because deviation from this norm was likely to be too costly (ZHANG, J., 2004).

In this model, all cells on the lattice grid were occupied meaning there were no vacancies. However, individuals were engaged in a process of switching locations based on the utility of the two cells. Utility or payoff was given to be a function of the number of like neighbours in surrounding cells, and the value of neighbourhood characteristics. During the simulation, randomly selected agent pairs swapped cells on the basis of maximising utility. The latter calculated by a series of mathematical functions where utility increased as the number of like surrounding neighbours increased. This continued until there were 50% agents of the same kind after which point the utility decreased. Agents were not required to swap cells if the payoff was not attractive. The model did not make use of real world data.

Aguilera and Ugalde (2007) attached house prices to each space on a two-dimensional lattice grid. The model was used to examine the relationship between income inequality and residential segregation. There were no vacant cells on the grid space. Instead, much like the model by Zhang (2004) agents swap places. Agents were governed by one behavioural rule – they moved to houses by matching their social status with the price of the house. House prices were updated systematically using a predefined control parameter. The results of the model supported Massey's (1996) theory which suggest that the more unequal the distribution of social class appears in neighbourhood, the more segregated it will become.

Benenson (2004) focused on residential dissonance and its relationship with residential segregation. Citing from the work of Goodman (1981) and DaVanzo (1981), residential dissonance was noted to be a tradeoff between the costs and benefits of moving from the old home and finding a new one. Based on the city of Yaffo, Tel-Aviv, the model attempted to recreate the behaviour of Jewish and Arab householders. The model was calibrated with real data set in an actual geography; using GIS data of streets and houses. Household data was based on census outputs and included information on age, education, ethnicity, marital status and income. House attributes included information such as the number of rooms in a house, household appliances and travel to work times among others.

Residential dissonance was derived from basic preference rules. These preferences alluded to the likelihood that Arab and Jewish households would occupy block houses, houses of oriental architecture, newly built block houses and homogeneous neighbourhoods. Arabs were thought to strongly avoid block houses and oriental architecture, Jewish householders preferred newly built block houses while both household types avoided homogeneous neighbourhoods concentrated with households of the other type. In general, Jewish households tended to avoid Arab households.

When agents were identified for relocation, 10 vacant houses were selected and the level of dissonance calculated based on the neighbourhood features at the new location. Once this list was generated and sorted, agents attempted to occupy the most attractive house, graduating through the list of houses until a house could be occupied. The results of the model illustrated that it was only sufficient that one group, either Jewish or Arab householders, had strong preferences to live among like neighbours as opposed to the theoretical view that segregation was only possible if both types of households exercised strong preferences.

Each of the models discussed shares some commonalities; building on Schelling's work, the models adapt, improve and/or amend Schelling's model in some way. Most of the models focus on behavioural rules based on income, social class and ethnicity while others primarily focus on the relationship between house prices and social class. Though some of the models are not applied to a real geography using real world data, they are still able to generate useful results which complement the theoretical literature. However, note that work by Benenson (2004) and Bruch and Mare (2006) use actual geographies with real world data. These models exhibited similar segregation trends. Most of the models are represented as closed systems; there are negligible updates to the model environment in the forms of population growth, for example. This is shown to have little impact on generating segregation trends. This is not to say that updating environmental variables within the model is not important, however.

ABM Models	**Geographic Data**	**Behavioural Rules**	**Calibration/Validation Methodology**	**Model Results**
Laurie *et al.* (2003)	50*50 edgeless torus of artificial agents; blacks to white ratio randomly determined	*R*-neighbourhood used to determine dissatisfaction in current and potential neighbourhood. *R* could be substituted for any value.	No calibration or validation against real world data	Segregation increased as the *R*-neighbourhood increased.
Bruch and Mare (2006)	500*500 grid of hypothetical agents; 50% black, 50% white	Move to areas where there are others of the same race. Move to areas where there are others with similar incomes.	No calibration; Resultant trends compared to trends in the 1978 and 1992 Detroit Area Study using absolute proportions	Residential mobility behaviour based on income decreased segregation while mobility behaviour based on race increase segregation.
Zhang (2004)	N*N lattice torus; ratio of agents vary but representative of two ethnicity types	Move to cell where the utility gained is more favourable than previous location.	No calibration or validation against real world data	Residential segregation could persist in a society even if individuals prefer to live in integrated neighbourhoods.
Aguilera and Ugalde (2007)	2D lattice grid where size = 1 to n; n*n agents	Move to cell where house price matched social class.	No calibration; Validated using Massey's segregation versus inequality theory and using Theil's inequality index	Segregation increases as inequalities across a neighbourhood increase.
Benenson (2004)	Israeli Census of Population and Housing 1995; ~40,000 individuals; GIS coverages of streets and houses	Arabs avoided block houses and oriental architecture. Jewish householders preferred newly built block houses.	Calibrated by changing the coefficient related to non-correspondence in the validation dataset and by altering agent attitudes of Arabs and Jews toward unfamiliar housing types and	Segregation could still occur if only one household group had strong preferences to live in neighbourhoods with like residents.

		Arabs and Jews avoided neighbourhoods primarily comprising of households of the other type.	neighbourhoods Validated against proportion of Arab/Jewish population; level of segregation using the Moran index; variation of the population when compared to the architectural style of buildings; annual fraction of households leaving Yaffo	

Table 3.2 Summary of agent-based models of residential mobility

All of the models illustrate how these factors influence segregation patterns in some way meaningfully contributing to the body of knowledge in the extant literature. These ABMs are used to unpack the dynamics influential in creating residential mobility behaviour in a way not possible in reality. The results are used to further describe the effects of the dynamics and the extent of their impact. Thus ABMs can be used to recreate residential mobility behaviour similar to the behaviour observed in reality. Details of these five models are summarised in **Table 3.2** above.

Even when other existing models are observed, these same trends are realised; limited behaviours are used to recreate residential mobility behaviours and few models use real world data applied to geometric spaces to analyse residential mobility behaviour. For example, research by O'Sullivan *et al.* (2003) and Fossett and Waren (2005) explored the influence of different neighbourhood sizes. This was similar to the work of Laurie *et al.* (2003). Pancs and Vriend (2007) examined the effect of preferences on integration policy. The results were similar to the work of Zhang (2004); even when individuals preferred integrated neighbourhoods, the impact of preferences led to segregated communities. Fosset and Senft (2004) increased the dynamics of their model to include preferences for a specific neighbourhood status and housing quality. Sethi and Somanathan (2004) explored the effects of preferences when the racial mix and average income were considered.

Much has been discussed in **Section 2.2** on the complexity of the residential mobility process. Unlike the models summarised in **Table 3.2** above, residential mobility is noted to be a function of a large number of interacting behaviours. These behaviours are based on factors such as life course triggers (births, deaths), income, social class, transport needs as it relates to employment, location of schools and exogenous factors – principally house price changes among others. It is apparent that the models listed above reduce the complexity of the residential mobility dynamics significantly, using at most three behavioural rules. Though apparently successful in creating segregation patterns, it is hypothesised that with increased dynamics, better abstractions of reality can be created using ABM. In this way, Schelling's model can be further extended to mimic an increased number of real world dynamics.

Extending the dynamics of Schelling's model is a contentious issue, however. Aguilera and Ugalde (2007) comment that dynamics must be sufficiently realistic and congruent with empirical research, thus suggesting that increased dynamics could lead to more realistic systems. On the other hand, Holden (1996) cautioned that by increasing the dynamics of the model it may be more difficult to understand causal relationships which may result. Such rationale should not be used as a deterrent when building more realistic models of residential mobility are concerned though it is important that individual level behavioural rules are clearly defined and verified to ensure correct execution.

In addition to this, many of the mentioned models do not make use of real world data in actual geographies. In the case of the models created by Benenson (2004) and Bruch and Mare (2006), behavioural rules were developed specifically for the study areas; Yaffo, Tel-Aviv and Los Angeles, California respectively, and all the input data, calibration and validation techniques used were specifically based on the study areas. But this is not true for the other more general models. In addition to extending Schelling's model to include more real world dynamics, real data in actual geographies can be used in model processing. As a result, calibration and validation procedures can be applied so as to test the validity of model results.

Coupled with the fact that existing residential mobility models are limited in dynamics and seldom make use of real world data, there are no spatially explicit agent-based models of residential mobility designed for the general UK housing market and calibrated/validated using appropriate real world data. Building such a model is a novel venture. Utilising real world data based on the EASEL district of Leeds can further assist in understanding the residential mobility dynamics in this area while allowing for the assessment of intended urban regeneration policy schemes.

3.4 An Introduction to a new ABM Approach to Modelling Residential Mobility

The new ABM of residential mobility is hereafter referred to as the CHAIRS model or the CHAIRS simulation. CHAIRS offers a new approach to modelling residential mobility by increasing the number of dynamics governing each household agent. The model uses actual spatial data in a geometric space representative of the EASEL district. Behavioural rules in the model are customised to the UK context though the model is calibrated and validated using EASEL-specific data. Resultant output from the model is then used to assess possible outcomes of urban regeneration schemes proposed for the case study area. Each of these features adds to the uniqueness of the CHAIRS simulation. This section is used to present these distinctive features; it is an introduction to the methodological framework that follows in **Chapter 4**.

3.4.1 Augmenting the Model Dynamics

The CHAIRS simulation extends the world of typical Schelling-like models by adding additional behavioural realism. This is encapsulated in rules which are derived from the extant qualitative literature discussed in **Section 2.2**. The literature suggests that the family life cycle plays an integral role in the residential mobility decision (DIELEMAN, F., 2001). The primary driver of the family life cycle is age. As individuals move from one stage of the life cycle to another life events such as births and deaths affect the composition of the household and may require larger or smaller homes with varied amenities (DIELEMAN, F., 2001). Likewise, as dependent children move from their teenage years to adulthood they are likely to leave the family home creating households of their own (CLARK, W.A.V. and Onaka, J. L., 1983). As mentioned in **Section 2.2**, these are the drivers which trigger the decision to move, part of the residential mobility process.

The household SAR encapsulates similar family life cycle events into one variable which is called the '*propensity to move*'. The '*propensity to move*' statistic records the probability that a household will move house over the subsequent year taking into account the age of the household, social class, educational achievement, housing tenure, employment status and health (STILLWELL, J. *et al*., 2008). The variable is used to determine whether a household will move house or not in a given year and is discussed extensively in **Section 4.3.3.1**.

Recall that residential mobility is a combination of two processes; the decision to move and a subsequent process of choosing a new house. This latter process is the relocation phase and is driven by the behavioural rules. **Section 2.2** highlights factors weighed by households when choosing a new home. These factors include the financial budget, familiarity with the neighbourhood, size of house required, racial tolerances, amenities in the local area (shops, medical facilities), access to educational institutions, access to transport routes for the purpose of employment and neighbourhood quality (**Section 2.2.2**). These factors are further influenced by housing market factors such as tenure structures, availability of suitable housing, changing house prices and mortgage interest rates, inflation and demographic changes across communities (**Section 2.4**). The behavioural rules are derived from a combination of these factors. The choice of rules is a balance between improving upon the dynamics of existing residential mobility ABMs and identifying parameters in real world datasets which appropriately quantify soft variables.

Thus far, augmentations of Schelling's model include dynamics which chiefly govern income, neighbourhood size and racial preferences (**Section 3.3**). The CHAIRS simulation will improve on these dynamics by including seven behavioural rules. The rules are noted in **Table 3.3** below:

	Behavioural Rule	**Supporting Literature**
1	Households move to areas to which they are familiar	(RABE, B. and Taylor, M. P., 2010); (BRIDGE, G., 2002)
2	Households move to better quality areas	(TU, Y. and Goldfinch, J., 1996)
3	Households move to houses where the number of rooms is satisfactory	(DIELEMAN, F., 2001); (LONG, L. H., 1972)
4	Households move to areas where the ethnic makeup is tolerable	(BRUCH, E. E. and Mare, R. D., 2006); (SCHELLING, T. C., 1969)
5	Households move to areas where transport	(RABE, B. and Taylor, M. P., 2010);

	routes are readily available	(GJESSING, M., 2010)
6	Households move to houses they can afford	(DIELEMAN, F., 2001)
7	Households containing school-aged children move to areas where schools are accessible	(LONG, L. H., 1972); (CROFT, J., 2004)

Table 3.3 A list of the behavioural rules to be used in the new agent-based simulation with supporting literature

There is much to suggest that rules 4 and 6 have been recreated in ABMs of residential mobility, however, this is not true for the other rules based on the literature previously discussed in this chapter. There is literature which suggests that households prefer to live in areas with which they are familiar (RABE, B. and Taylor, M. P., 2010) and in line with the hierarchy of housing tenures discussed in **Section 2.2.1**, the overall goal for each household is to live in a better house in a better neighbourhood therefore moving to improvement. In the case of the chosen rules, a better house is defined as a house with a satisfactory number of rooms; avoiding issues such as overcrowding. Like the Schelling model, the ethnicity rule is repeated and it is implemented in a manner similar to Bruch and Mare (2006) where both the ethnic makeup of the immediate neighbourhood and the potential future neighbourhood are considered. Affordability is captured by rules 5 and 6. Rule 5 alludes to the fact that accessible transport routes are needed so as not to hamper the journey to work while rule 6 uses social class as a derivative of income. The final rule attempts to relocate households with school-aged children in close proximity to schools.

Note that there are additional rules detailed in the literature which have not been included in this model. For example, some households prefer to live in areas where amenities such as grocers and doctors' surgeries are easily accessible (RABE, B. and Taylor, M. P., 2010). Likewise, some households prefer to live in areas where they can maintain their social networks, that is, the desire to live close to family and friends (BRIDGE, G., 2002). Such behavioural rules were not included in this model due to the lack of available data. Also to some extent, the list of housing choice behavioural rules is not exhaustive, however, the chosen rules are thought to be a reasonable balance between extending Schelling's model and utilising suitable parameters in the extant qualitative literature.

These behavioural rules provide a mechanism for households to interact with the surrounding environment; neighbourhood quality is assessed using the Index of Multiple Deprivation (IMD) (**Section 4.3.3.2**), distances from schools and transport routes are generated using the geometrically coordinated data linked to both of these entities. The outcome of analysing each rule is weighted in such a way that only the best house available is assigned to the household engaged in the relocation process (**Section 4.3.3.2**). There are no updates made throughout the lifetime of the model with regards to population growth, in/out-migration, aging, or changes in the physical landscape originally used. This is presented in the discussion on model assumptions, **Section 4.5**.

3.4.2 Using Spatial Data in Realistic Geometric Spaces

The model utilises survey and GIS data in order to generate results. All data is based on the EASEL district only and includes aggregate 2001 Census statistics by OA, individual level records from the 2001 household SAR, survey data from Acxiom's ROP and geometric data such as OAs, schools, houses and roads stored in ESRI (Economic Social Research Institute) shapefiles. Note that OAs are the smallest census geography available, there are generally comprised of approximately 125 households (OFFICE OF NATIONAL STATISTICS, 2011a) while ESRI shapefiles are coordinated database files used to represent spatial objects pictorially (ECONOMIC AND SOCIAL RESEARCH INSTITUTE, 2011).

Both the 2001 Census and 2001 household SAR are used to create individual level records representative of households in the EASEL area though a process of microsimulation. In addition to shapefile data, household records are added to the simulation. Households are assigned to houses and use a combination of OA, roads and schools geometric and aspatial data in making the residential relocation decision. This is discussed in detailed in **Sections 4.1** and **4.2**.

3.4.3 Applying Calibration and Validation Techniques with EASEL-Specific Survey Data

Calibration and validation are procedures used to examine the results of computer models. The process of calibration is used to ensure that parameters in the model generate behaviours similar to those observed in the real world (DE SMITH, M. J. *et al*., 2009). Validation, on the other hand, ensures that the model recreates reality with a satisfactory measure of accuracy ((SARGENT, T. J., 1998) in (BIANCHI, C. *et al*., 2008)). Calibration and validation procedures are often found in economic models whose basis is of a mathematical nature. Such procedures, however, can be used in examining the results of microsimulation systems and other computer models (DAWKINS, C. *et al*., 2001) and help to bring realism to model abstractions by comparing and constraining model outputs with real world data. Thus economic models may use stock market details in an attempt to predict future trends in the market while demographers may use census data in microsimulation models to estimate population growth over time. In both of these examples, model results can be tested against live data as time progresses.

To date, calibration and validation procedures are not commonly used when examining the results of agent-based systems many of which are very abstract. For example, Schelling did not use actual data in his model neither were there allowances made for calibration or validation. That does not mean that Schelling's contribution to the ABM literature must be ignored, having established the rudiments of segregation theory in the form of a model.

In general, agent-based systems are created and used to unpack and understand the cause of complex behaviour at the aggregate level by focusing on behaviour at the individual level. Thus, it is the evolution of simple individual level behaviour into complex aggregate behaviour that is of interest. Keeping track of individual agents is a challenge, however and oftentimes aggregation obscures spatial variation. Considering that existing datasets likely to be used for calibration/validation are often presented at the aggregate level, calibrating/validating agent-based models is a challenging activity. Coupled with this, though an ABM may be able to produce countless amounts of variable output throughout the lifetime of the model, data available for calibration/validation may not be reported with the same level of detail (CROOKS, A *et al*., 2008). For this reason, not only do modellers admit that calibrating/validation ABMs is difficult, some modellers claim that these processes are not possible (WEBSTER, F. V. *et al*., 1988). Despite this, if ABM theory is to be extended into a real world policy making tool, attempts must be made to applied calibration/validation techniques so as to increase the realism of model outputs.

One unique element of this model lies in the use of the ROP to calibrate/validate simulation results. This is an individual level dataset comprising of data from households in the EASEL district. **Chapter 5** discusses this dataset in more detail, illustrating how calibration and validation techniques can be applied to the CHAIRS simulation. Using realistic data in analysing the validity of model outputs in this way is significant to ABM literature primarily because this is not yet a routine part of assessing ABM outcomes particularly in models founded on the Schelling system.

3.4.4 Assessing Policy Implications Using 'What-if' Scenarios

Scenarios are defined and used to assess policy implications as a result of behaviours in the CHAIRS simulation. By so doing, real world outcomes can be analysed in a way not possible in reality; policy implications take time to observe in reality while simulation models can produce results in a matter of minutes. Once the CHAIRS simulation is constructed, calibrated and validated to ensure realistic outcomes, policy focused scenarios can be implemented and explored, providing a mechanism for analysing possible future trends. More specifically, the CHAIRS simulation is used to analyse the likely outcomes if two housing-led regeneration schemes are applied; building a mixed-tenure housing development and changing the transport network. These outcomes are presented in **Chapter 6**.

3.5 Conclusion

In this chapter, computer modelling is discussed as the quantitative technique to be used in simulating residential mobility behaviour and the implementation of urban regeneration policy. Several modelling techniques were identified; spatial interaction modelling, microsimulation modelling, CA and ABM. ABM was selected as the modelling technique of choice because of its ability to represent dynamic behaviour at the individual level, a characteristic not entirely possible through the use of other modelling techniques. Several existing models of residential mobility were presented. Each of the models was used to illustrate the small number of behavioural rules used to simulate residential mobility behaviour noting that the residential mobility process is a combination of several more behaviours which to date have not been modelled extensively. For this reason, the CHAIRS simulation is introduced. It is a model of residential mobility with an increased number of behavioural rules to govern the mobility process. The methodological framework driving this model is now presented in the chapter to follow, **Chapter 4**.

Chapter 4
An Introduction to the Methodological Framework for the CHAIRS Simulation Model

4.1 Introduction

Thus far, the dimensions of the research question have been explored with a detailed examination of the extant literature. **Chapter 2** discussed the nature of the housing market, housing policy and housing choice while **Chapter 3** explored to the usefulness of computational modelling as a means of gauging answers to research questions in the realm of human/social geography. This chapter is used to introduce and discuss the methodological framework for the proposed model. It outlines the core algorithmic and processing details of the residential mobility simulation model CHAIRS. As introduced in **Chapter 1**, the acronym CHAIRS is used to describe the goal of the housing-led regeneration plans in the EASEL district, that is, creating housing alternatives.

The chapter is structured in such a way that all inputs, processes and outputs of the model are discussed in relative detail. This tripartite structure is general for most computer programmes; **Figure 4.1** is used to illustrate how these three elements flow in the context of the simulation model. The term '*inputs*' is used to describe any initial data needed to facilitate the correct execution of the model. The processing stage allows for the manipulation and analysis of any input data used; it is this stage at which complexity may be generated and is a stage where some of the interaction between processes may not be clearly understood, as discussed in **Section 4.3**. Outputs are the results of model processing. Initial outputs will be described, though a more detailed analysis of the outputs will be presented in **Chapters 5** and **6** where the calibration/validation processes and final results are presented respectively.

Figure 4.1 Model Development Process

Written in the Java programming language, the CHAIRS model has been built in the Repast Simphony environment. The Repast Simphony Integrated Development Environment (IDE) is a programming application which relies on a mixture of the core java classes and integrated geographic classes (MACAL, C. M. and North, M. J., 2010). Repast is specifically designed for developers with an interest in the creation of agent-based models. This is different to pure java IDEs such as NetBeans which are not specifically designed for ABM though plug-in libraries tools such as Geotools can be used to simulate such. Repast is the successor of basic ABM programmes such as NetLogo and StarLogo.

All model functionality was first built and tested on a PC in the Repast environment using a subset of the input data, as due to the limitations in PC processing capabilities, full execution of the model was only possible when using more powerful computing facilities. For full model runs the Advanced Research Computing (ARC1) supercomputing environment resident at the University of Leeds was used. ARC1 is a Linux-based supercomputing environment with 4000 CPUs, 7Tb memory and 1Tb of disk space. By using both the Repast and ARC1 applications, the CHAIRS simulation model could be explored extensively. Though the real time visualisation feature available on the PC version of Repast is not yet available when the simulation is scaled up and executed in the supercomputing environment.

Such a model can be defined in terms of the major algorithmic detail and this is described below. The model simulates the interaction of households trying to find a new place of residence and features relationships between the environmental variables such as OAs, Roads and Schools. Households are the agents in this model; continually interacting with the environment around them in order to find a new house of residence or housing unit. The basic algorithm used to implement the model is also accompanied by a flowchart in **Figure 4.2** which is used to give a pictorial view of the programmatic logic. The algorithm is as follows:

1. **Read** in all input data (*houses*, *households*, *output areas*, *schools*, *roads*)
2. **Initialise** data structures; **set** house attributes, **assign** *households* to *houses* etc. (**Section 4.3.1**)
3. Begin simulation time step counter (**Section 4.3.2**)
4. **Repeat for** each *household*
    a. Assess the 'propensity to move' (**Section 4.3.3.1**)
    b. **If** *household* chooses to move **Then** (**Section 4.3.3.2**)
        i. Select a subset of *houses*
        ii. Find a new *house* from this subset based on the behavioural rules
        iii. Move to most attractive house
    c. **If** 1 year has elapsed **Then**
        i. **Print** *house* to *household* allocation for year x (**Section 4.4**)
5. **Repeat** from step 4 until user stops programme

The texts which have been italicised, emboldened and underlined are significant. The italicised text highlight the major entities used in the model (houses, households etc.). The emboldened text highlight the programmatic constructs while the text which have been underlined show some of the variables of interest which will be discussed in this chapter. The algorithm is further depicted in the flowchart figure below, **Figure 4.2**.

Each of these elements will be further discussed in this chapter. Once started, the model can be terminated at any time by the user using the Linux-based command line interface controlling the ARC1 model execution, however, a short batch file is also used to terminate each model run after a period of two hours has elapsed. The details presented in this chapter will follow the flow presented in the **Figure 4.2**. The input data used in the CHAIRS model will be discussed first, the processes used to drive the system are then presented and finally, a basic idea of the output generated is outlined. It should be noted that household data is created by the process of microsimulation. This is discussed in detail in **Section 4.2.2**. The microsimulation model is only executed once in order to generate the synthetic population as at 2001. The synthetic population is created before the full execution of the CHAIRS simulation.

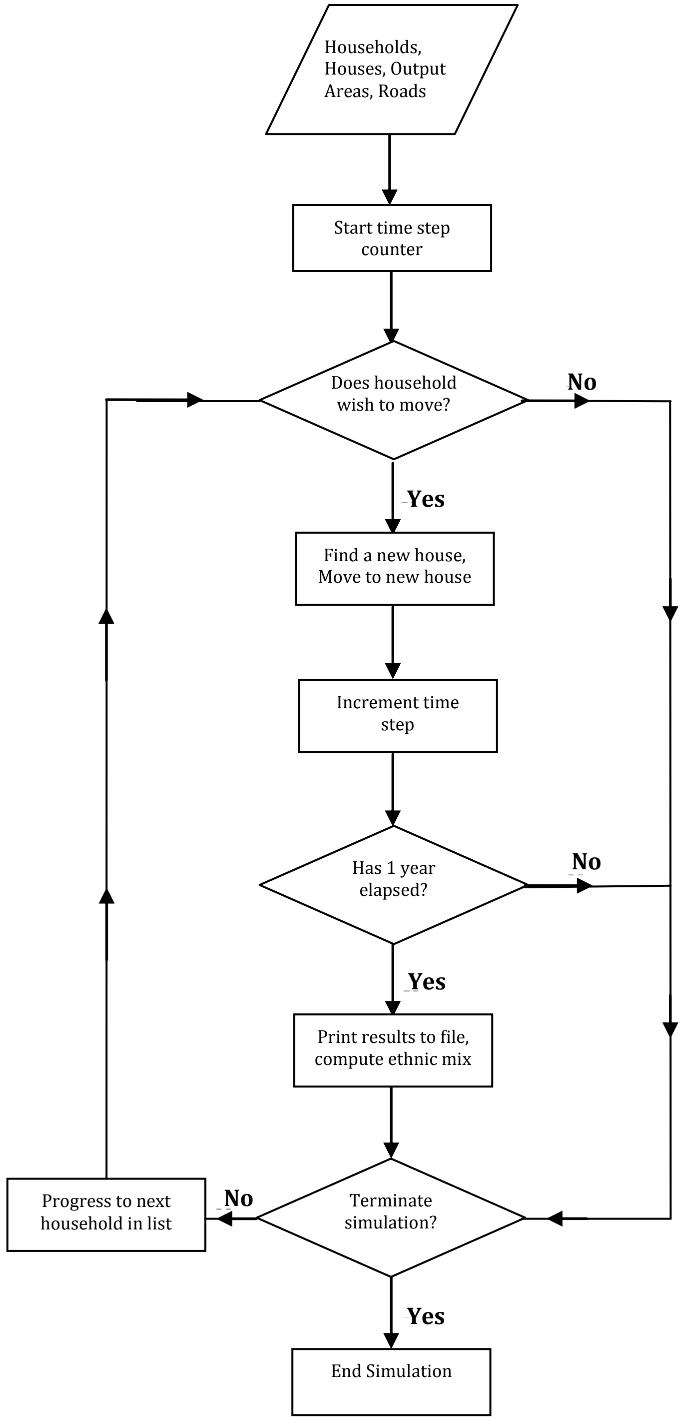


Figure 4.2 Flow Diagram describing the CHAIRS simulation

4.2 Model Inputs

There are five categories of input data used in the CHAIRS simulation. These inputs comprise of data on households, houses, OAs, schools and roads. Recall that OAs are census areas with approximately 125 households (**Section 3.4.2**). Each of these entities is used to represent the EASEL area by creating features similar to the real world environment. For a residential mobility model such as this, these features are important in the context of the behaviours chosen to simulate residential mobility (**Section 2.2**). Households move to and from various houses and the type of neighbourhoods chosen is very important. Neighbourhoods are loosely represented by OAs. Though other neighbourhood categorisation models such as community areas could be adopted, OAs have been selected so as to allow for the use of IMD data which is available at the OA level. Arguably, neighbourhood categorisation models such as community areas take into account the natural dimensions and dynamics of real communities (REES, P. *et al.*, 2004), however, the IMD data available at this level is not readily available. Features such as schools are also useful for households with dependent children when deciding to find a new home, while roads help when transport decisions are made. These features are detailed in **Table 4.1** below.

Feature	**Data Type (Shapefiles)**	**Source**	**Attributes of interest (parameters)**
Houses	Polygon	Ordnance Survey Mastermap	Contains house Ids only
Schools	Point	Education Leeds	School type; primary or secondary, school name
Output Areas	Polygon	Edina Digimap	IMD, OA code
Roads	Line	Ordnance Survey Mastermap	Road classification; major road or pedestrian walkway

Table 4.1 List of shapefile features used in the model with details on the data types and sources

Households are noted to be the agents, it is assumed that each household resides in a house. This therefore ignores institutional populations. Houses are located within OAs where roads and schools exist. Each household interacts with this environment when making the residential relocation decision.

4.2.1 Input Data Derived from Shapefiles; houses, OAs, schools, roads

All data related to houses, OAs, schools and roads are derived from ESRI shapefiles. Note that a shapefile is a specially coordinated database file used to represent spatial objects pictorially (**Section 3.4.2**). Such objects may take the form of points, lines and polygons. In the CHAIRS simulation, houses and OAs are stored in polygon shapefiles, schools in a point shapefile while roads are stored in a line shapefile as previously noted in **Table 4.1**.

All OA data have been downloaded from the Edina Digimap online source, data on schools have been obtained through Education Leeds, while shapefiles representing houses and roads were obtained from the Ordnance Survey. In addition to the coordinate data contained in each shapefile used, the shapefile containing OAs also stores statistics representative of the IMD. Indices such as the IMD are used to rank OAs according to their relative level of deprivation and use indicators pertaining to income, employment, health deprivation and disability, education skills and training, barriers to housing and services, crime and living environment (LEEDS CITY COUNCIL, 2007c). The shapefile detailing the location of roads contains information on road classification, namely, '*car walk majorRoad*', '*car majorRoad*', '*car walk*' or '*walk*', where major road indicates a primary road, '*car walk*' refers to a secondary road and '*walk*' refers to a pedestrian pathway.

The shapefiles containing information on houses and schools respectively only contain geographical references to the shape object without any additional data. In this case, there is no distinction between different house types or school types. There are 9 schools, 286 OAs and 38139 houses. How these objects are used will be discussed in the context of model processing (**Section 4.3**), however, the map below (**Figure 4.3**) is a pictorial view of the four shapefile object types. All input data remains static through the model processing phase, that is, there is no feedback in the system to update data such as IMDs, number of schools etc.

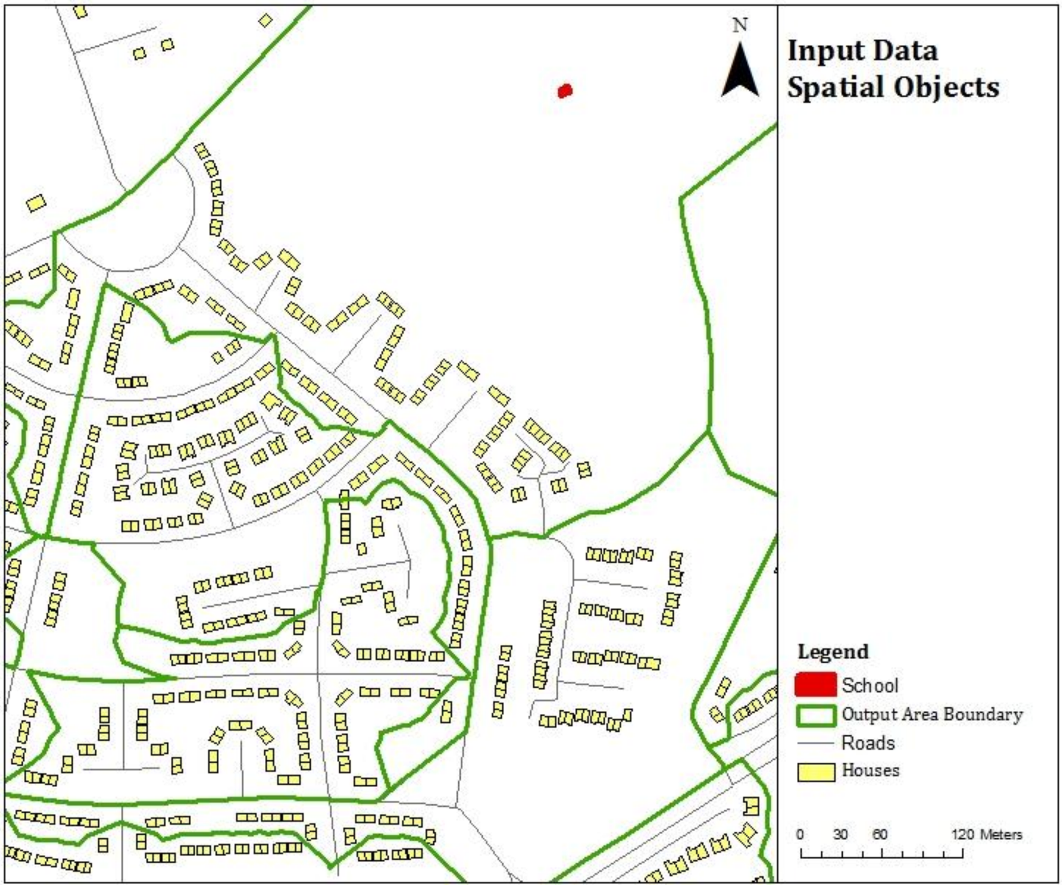


Figure 4.3 Pictorial Illustration of Input Data Derived from Shapefiles

4.2.2 Input Data Derived by Other Means; households

Obtaining data related to households requires a more involved process of data manipulation as there is no one dataset, readily available in the public domain, and containing individual level records of households or individuals representative of the EASEL area. Such a process includes gathering information from the 2001 Census and the 2001 Special Licence household SAR as well as using the modelling technique of microsimulation to create a full dataset of EASEL area residents. In 2001 a total of 35729 households were recorded in the EASEL district. The microsimulation process will be used to create a dataset with the same number of households to be used as the input for the CHAIRS simulation.

4.2.2.1 The Population Datasets; Household SAR, 2001 UK Census

The 2001 household SAR is a 1% sample of anonymous, geography neutral, individual records from the 2001 UK Census (CENTRE FOR CENSUS AND SURVEY RESEARCH, 2011). This dataset contains approximately 520,000 records each representative of one individual from various locations in the United Kingdom. The dataset can also be aggregated into households using the '*household representative person*' as the aggregate key. The '*household representative person*' is used to represent the head of the household. When aggregated, the household SAR contains approximately 225,000 households and contains information such as age, gender, ethnicity and marital status at the individual level, and accommodation type and tenure type at the household level. This dataset is appropriate as it contains demographic details, it can be used to recreate a dataset representative of the EASEL area.

Unlike the household SAR, the 2001 Census does not contain individual records. It is an aggregated dataset and contains information on each OA in the UK. Such information includes key statistics on age structure, population counts, socio economic status, tenure types, qualifications, and household structures. Cross-tabulated data are also available to provide different dimensions at which the population can be analysed. For example, tables containing information on 'Age by Social Class', 'Age by Tenure by Social Class' and 'Social Class by Tenure'. Aggregate data directly related to the EASEL district is contained within the 2001 Census. By using this dataset in collaboration with the household SAR, a full synthetic dataset of EASEL area households can be created.

In the context of this project, the household SAR is used as the sample dataset containing household records which may or may not be like those households in the EASEL area. The 2001 UK Census contains population counts for different census variables in the EASEL area. A table of the Lower Level Super Output Area (LLSOA) is used to represent the EASEL. The LLSOA is the smallest census geography, each LLSOA contains on average 125 households (**Section 3.4.2**). By using the technique of microsimulation, records from the household SAR can be selected based on some condition. These records can then be used to create the EASEL population, guided by the 2001 Census counts.

4.2.2.2 Using the Microsimulation Model to Create the Synthetic EASEL Population

The microsimulation model used is a component of the Flexible Modelling Framework (FMF) created by Harland and described in Smith *et al.* (2009). In essence, the microsimulation examines records in the household SAR and chooses the household records with profiles that best represent the residents of the EASEL district. The model is deterministic in nature as it matches individuals from the sample population found in the household SAR to a specified geography from a census-defined population of the 2001 Census by re-weighting individual records based on chosen census variable constraints (SMITH, D. et al., 2009). In general, synthetic population estimation is commonly understood and not a major objective of this project.

The chosen constraints are used to skew the synthetic based population in such a way that the significant features are represented. For example, though the household SAR contains a general population of several different types of households, the EASEL district contains a significant number of social housing tenants in some districts. Thus, if tenure type is chosen as a constraint, during the microsimulation process and Area **A** has 100 social housing homes, then the microsimulation model will choose 100 social housing households from the sample population and place them in the geography of Area **A**. These records are then appended to the table storing the synthetic population. In a similar way, if tenure and socio economic status are the chosen constraints, during the microsimulation exercise, if Area **B** has 60 social housing tenants while 50 of the households in those homes are employed in routine occupations, then the model will choose households from the sample population with these specific profiles and place them in the geography of Area **B**. The choice of constraints is therefore important, as constraints determine how well the synthetic population match the actual population. More constraints will improve the match with the underlying 'real' population, however, the match will nevertheless rarely be perfect as increasing constraints makes it harder to find SAR individuals to match.

Choosing the Constraint Variables

Based on the area profile presented in **Section 1.3**, the EASEL district is comprised of some of the most deprived districts in England. It is an area where there is a high concentration of social housing tenants including ethnic minority groups thought to be socially disadvantaged. Choosing appropriate constraints for the microsimulation process is therefore important so as to reproduce a synthetic population that mirror these significant features. Recall that constraints ensure that each record chosen from the sample population, the household SAR, is representative of a household in a specific area of the EASEL district. They are used to skew the synthetic population in such a way that it matches the EASEL population. Noting that indicators of deprivation are important in recreating the individual level EASEL population, variables such as social status, accommodation type, number of cars and economic activity are some of the indicators that may be chosen as constraints (PALMER, G. *et al.*, 2008). Therefore the choice of constraints is a combination of understanding the significant features of the EASEL area and testing the variables systematically.

Regression analysis allows for a systematic process of choosing constraint variables however. This statistical technique is used to assess the relationship between independent variables and some dependent variable (HARRELL, F. E., 2001). For independent variables, variation does not depend on any other variable while the converse is true for dependent variables (HARRELL, F. E., 2001). Using the household SAR, the '*propensity to move*' statistic is the dependent variable as it indicates the likelihood that a household would move house. Using SPSS, **Table 4.1** is a list of independent variables that will be used in the regression model. It should be noted that due to the large number of variables in the household SAR, these variables were selected based on information presented in the literature (**Section 2.2**) as a first sweep attempt to reduce the number of variables used. In this way, the incidence of co-linearity was further reduced. A more detailed description of the variables can be found in **Appendix A**. Due to the large number of categorical variable contain within the household SAR, binary logistic regression is the regression technique of choice. Note that for this reason, linear regression is not appropriate (HOSMER, D. W. and Lemeshow, S., 2000).

Variables	Brief Description
Accommodation type	Semi-detached, detached etc.
Family type	Lone parent, Married with children etc.
Social Status	Skilled, Unskilled, etc.
Tenure	Private owned, Private rented, Council rented
Number of cars	0, 1, 2 or more
Number of rooms in occupied house	0 – x
Ethnicity	White, Irish, Chinese, Pakistani etc.
Household Density	0 – x

Table 4.2 List of Potential Constraint Variables

It is important to note that each record generated by the microsimulation will be representative of one household. This is based on the assumption that when a household chooses to move, everyone in the household is affected. Therefore, though other potential constraint variables could be added to the list presented above, the fact that these variables are not recorded at the household level in both the household SAR and the 2001 Census disqualifies them from this process. Such variables include, for example, age, employment status, country of birth, religion and educational qualifications; the census does not record this data at the household level. The parallels between the household SAR and the 2001 Census are necessary as both datasets are used in creating the new synthetic population.

Using binary logistic regression, all variables used were found to explain the '*propensity to move*' in some way. Here the '*propensity to move*' was regarded as the dependent variable comprising of the categories moved/not moved. Each of the independent variables in **Table 4.2** was used to test for a significant relationship using binary logistic regression in SPSS. The final choice of constraint variables is based on a combination of what is known about the EASEL district in collaboration with the results of this regression exercise. **Table 4.3** below shows the results of the regression model using the variables outlined in **Table 4.2**.

Variable	Category	Wald Statistic	Significance	Likelihood of Migrating
Tenure	Owners **(ref)**	5681.154	.000	
	Public renters	206.900	.000	1.397 (1.335-1.462)
	Private renters	5462.442	.000	4.288 (4.125-4.456)
Family Type	Lone parent **(ref)**	4045.842	.000	
	Married couples	467.800	.000	.572 (.544-.602)
	Cohabiting couples	740.510	.000	2.122 (2.01-2.24)
	Ungrouped individuals	*1.509*	*.219*	*.969 (.921-1.019)*
Social Status	Higher and Lower Managerial Professionals **(ref)**	343.981	.000	
	Intermediate Occupations, Small employers, Lower Supervisory	165.203	.000	.763 (.732-.795)
	Semi-routine and routine occupations	281.369	.000	.665 (.634-.698)
	Never worked, long-term ill, not classified	255.257	.000	.642 (.608-.678)
Accommodation Type	Detached, Semi-detached **(ref)**	300.789	.000	
	Terrace	103.561	.000	1.199 (1.158-1.242)
	Flats, Temporary Accommodation	287.226	.000	1.438 (1.379-1.5)
Ethnicity	White **(ref)**	247.280	.000	
	Mixed Ethnicity	50.415	.000	1.617 (1.416-1.846)
	Asian	172.151	.000	1.636 (1.520-1.761)
	Black	*.022*	*.883*	*1.007 (.923-1.098)*
Number of Cars	No cars **(ref)**	68.267	.000	
	1 Car	*.018*	*.892*	*1.003 (.963-1.045)*
	2 or more cars	35.485	.000	1.162 (1.106-1.220)
Density	1+ Persons per Room	28.790	.000	1.284 (1.172-1.407)
Number of Rooms in Occupied House	2+ Rooms	*.005*	*.942*	*1.005 (.883-1.143)*

Table 4.3 Regression Model using variables defined in *Table 4.1*

In the table above there are three statistics presented; Wald, Significance and the Likelihood of Migrating. The Wald statistic is a chi-square statistic that tests the strength of the relationship between the variables under observation and the propensity to move statistic; the higher the value of chi-square the stronger the relationship (HOSMER, D. W. and Lemeshow, S., 2000). The Likelihood of Migrating variable explains the level of difference between individual categories of each categorical variable. If these values overlap then the overlapping categories are not significantly different in predicting the outcome variable, the propensity to move

(Harrell, 2001). The column labelled Significance shows the statistical significance of the variable under examination. It is benchmarked against a value of 0.05 which is noted to be the accepted standard used in most regression analyses (HOSMER, D. W. and Lemeshow, S., 2000). A significance value less than 0.05 indicates that the regression model is statistically significant while .000 indicates a very high significance. Note that the label **(ref)** is used to indicate the reference category where all statistics in the reference groups are made relative to the chosen reference category. For example, for the reference group '*number of cars*', '*no cars*' is the reference category.

Table 4.3 shows the results of the regression model. Variables in the table are in order of significance with the most significant variable being tenure. Here, the chi-square Wald statistic is noted to be higher than all other variables. This is an indication that the relationship between the tenure variable and the propensity to move statistic is stronger than the relationship between family type and propensity to move statistics when all variables in the model are considered. The predictive power of the family type variable is also defined. This is followed by social status, accommodation type, ethnicity, number of cars, density and number of rooms in occupied house. In terms of significance, variable categories such as ungrouped individuals, Black, 1 car and 2+ Rooms have significance values higher than 0.05. Here the significant values simply mean that there is no difference in the likelihood of migrating when compared to the reference category.

The variables which contribute most to the propensity to move statistic are the ones chosen initially. These include *tenure*, *social class*, *family type* and *accommodation type*. As documented by Smith *et al.* (2009), 4-5 constraint variables are sufficient in order to produce the intended synthetic population. The variables detailing *ethnicity*, *number of rooms in occupied house* and *household density* have not been chosen because of the lower ranking illustrated in the table above (**Table 4.2**). However, *number of cars* is chosen to improve the dimensions of the deprivation indicator needed in recreating the EASEL district, this variable allows for the creation of cross tabulated tables used in validating the synthetic population created through microsimulation.

Using the Constraints in the Microsimulation

With the constraints chosen, they must be prepared for use in the microsimulation model; for each constraint chosen, an equivalent table in the 2001 Census must be found at the OA level. In this way a synthetic population can be created to be used as the input population in the residential mobility model. All tables found in the census must be categorised in the same way as the variables in the households SAR, that is, if tenure is categorised into the three categories, owners, private renters, social renters then the tenure table taken from the census must be categorised in the same way. Secondly, cross-tabulated tables must be compiled in order to validate the aggregate values matched in the sample population and the constraint tables. The '*social status of household representative person by tenure*' and '*tenure by number of cars*' table are used for this evaluation procedure. Finally, a table containing the total household population for each EASEL OA must also be created. Counts of households from the constraint and evaluation tables are compared to ensure that they match proportionally with the total household population for each OA.

Once this data is prepared, it acts as the input to the microsimulation model. The model uses a process of reweighting whereby individual household records from the household SAR that match the constrained demographic characteristics from the 2001 Census are cloned until the population of each OA is created. The reweighting is repeated until each household in the household SAR is proportionally fitted to reflect the probability that that household would be in each OA of EASEL (SMITH, D. *et al.*, 2009). Thus every household in the household SAR has the opportunity to be allocated to every OA of EASEL, however, Smith *et al.*, (2009) notes, in some cases, no clones of individuals may appear while in other cases several copies of a single household may appear. **Table 4.4** is a sample of the output generated from the microsimulation. The duplication of household ids signifies the outcomes of the cloning process. When the microsimulation is complete, unique household ids are reassigned.

LLSOA	Household Id	Accommodation Type	Age	Tenure	Social Status	...	Economic Activity
00DAFF0001	1	5	3	1	1		1
00DAFF0001	1	5	3	1	1		1
00DAFF0001	2	3	6	1	2		1
...	...	...	...	...	...		...
00DAFF0001	3	1	4	1	2		1
00DAFF0001	4	3	4	1	4		4
00DAFF0001	4	3	4	1	4		4

Table 4.4 Sample Output of Microsimulation

Also note that though there were only five constraint variables used additional attributes such as age and country of birth appear in the sample output above. When individual households are identified for use, for example household id 1, all attributes associated with this household can be appended to create a full dataset of EASEL area households. There are 13 attributes that have been used in the CHAIRS simulation (**Table 4.5**). The use of these attributes will become clearer when model processing is presented. A more descriptive list can be found in **Appendix A**.

Variables	Brief Description
Household Id	Unique household identifier
Accommodation type	Semi-detached, detached etc.
Family type	Lone parent, Married with children etc.
Tenure	Private owned, Private rented, Council rented
Number of cars	0, 1, 2 or more
Number of dependents	0 – x
Social Class	Skilled, Unskilled, etc.
Age Category	16-25; 26-35 ... >75
Number of rooms in occupied house	0 – x
Number of rooms required	0 – x
Number of residents in occupied house	0 – x
Ethnicity	White, Irish, Chinese, Pakistani etc.
Output area code	Unique output area identifier

Table 4.5 Attribute Data for each household record used in the CHAIRS simulation

In addition to generating a list of households, by using the total number of household spaces noted in the 2001 Census, a full list of houses is also generated, vacant and occupied. This list, also derived from the census, stores accommodation types, housing tenure, number of rooms and OA data. Both the individual level household data and the individual level house data will be used in the model processing.

Two goodness of fit measures have been used to ensure that the results of the microsimulation are acceptable, R^2 and Square Root of the Mean Square Error (SRMSE). R^2 is used to measure how well the real data matches a regression line (KNUDSEN, D.C. and Fotheringham, A.S., 1986). An R^2 value of 1 indicate a perfect match to the real data points. In practical terms, the SRMSE measures the differences between the data generated by the microsimulation model and the actual observed values. This therefore means that an SRMSE value close to zero is ideal (KNUDSEN, D.C. and Fotheringham, A.S., 1986).

Table 4.6 below details the results goodness of fit statistics for the constraint variables. Here both the accommodation type and socioeconomic status variables appear to match the original data perfectly while household composition, tenure and number of cars are very close to the ideal outcomes. Take note that though individual level data is created, these goodness of fit measures assess the results at the aggregate OA level.

Variable Name	R^2	SRMSE
Accommodation Type	1	0
Household Composition	0.938	0.299
Socioeconomic Status	1	0
Tenure	0.955	0.233
Number of cars	0.944	0.233

Table 4.6 Error statistics R^2 and SRMSE used to illustrate the goodness of fit for the data generated through microsimulation

4.3 Discussing the Algorithmic Details of the Model (Processing)

Using the newly created synthetic population along with all other input data the actual processing details of the model can be discussed. The model process is re-introduced in the detailed steps below. Already described in the previous section (**Section 4.2**), **Step 1** allows for data to be read into the model. These inputs include houses, schools, roads, OAs and households. It is the steps that remain which will be discussed in this section of the chapter.

1. **Read** in all input data (*houses*, *households*, *output areas*, *schools*, *roads*)
2. **Initialise** data structures; **set** house attributes, **assign** *households* to *houses* etc. (**Section 4.3.1**)
3. Begin simulation time step counter (**Section 4.3.2**)
4. **Repeat for** each *household*
    a. Assess the 'propensity to move' (**Section 4.3.3.1**)
    b. **If** *household* chooses to move **Then** (**Section 4.3.3.2**)
        i. Select a subset of *houses*
        ii. Find a new *house* from this subset based on the behavioural rules
        iii. Move to most attractive house
    c. **If** 1 year has elapsed **Then**
        i. **Print** *house* to *household* allocation for year x (**Section 4.4**)
5. **Repeat** from step 4 until user stops programme

4.3.1 Initialise Data Structures

Each data file is read into the simulation and stored in separate arrays. At this stage, though each household has OA locations, they are not assigned to individual houses. In a similar way, though each house object has a geographical location, there are no attributes attached. Thus attributes must be assigned to housing units and households need to be assigned to houses. Noting that the microsimulation process generates a full list of households (occupied houses) and houses (vacant and occupied houses), the list of household's housing data is used to populate the attribute fields for each house record. Also by identifying houses by OA, for each OA the OA code is assigned to each house in the list from the OA dataset. During this process, households are also assigned to houses by matching the accommodation type, tenure, number of rooms and OA attributes. Once this is completed, the proportion of each ethnic group is calculated to be used in preparation for use in the behavioural rules. Finally, the initial household allocation is stored in a text file.

4.3.2 Begin Simulation; Start Time Step Counter

When the simulation begins the inbuilt Repast tick counter is used to measure time. This is a real time counter used in this simulation to measure the execution time of each simulation run. Time, as it relates to the simulation, however, is measured based on population movement. That is, according to the original statistics used in the household SAR, at least 14% of the population moved in the year 2001. Therefore, in this simulation a year elapses after 14% of the population has moved. Yearly time steps are important because as households are confronted with the residential mobility decision, this 14% threshold is used to limit the number of households that move in a given year. This is thought to be reasonable particularly because local area changes are more noticeable over longer time periods. For example, the fact that one British White household moves to a predominantly East Asian area is not thought to be significant until several British White households do the same. Such an event is not likely to be realised for some time, in this case, this time period is limited to 1 year.

Once the model has been initialised with the relative input data and the time counter started, the simulation can begin. In essence, each household is interrogated to establish whether the household would like to move then a new home is found until the user terminates the simulation.

4.3.3 Simulating the Residential Mobility Behaviour

Residential mobility behaviour can be simulated by determining if households wish to move and the process of finding a new house to move to. A decision tree construct is used to determine whether households wish to move (**Section 4.3.3.1**) while the behavioural rules are used to find a new housing unit for each household desirous of moving (**Section 4.3.3.2**).

4.3.3.1 Determining if household X should move

After initialising the model, the household list is traversed, examining households desirous of moving and finding new homes. As the household list is traversed, the '*propensity to move*' statistic is analysed (**Step 4a**). The '*propensity to move*' statistic is used to determine whether a household would choose to relocate or not. It is the probability that a household will move house based on household characteristics. If this statistic is less than a randomly generated number, this signifies the household's desire to move. The converse is true if the statistic is greater than the randomly generated number. Both the propensity to move statistic and the randomly generated number range from 0 to 1.

Deriving the '*propensity to move*' statistic

The '*propensity to move*' statistic is derived from the SAR, it is used to group households according to their probability to migrate. This probability is based on each household's characteristics. The statistic is derived from the household SAR using Chi-Squared Automatic Interaction Detector (CHAID) analysis. This is a chi-square analysis technique which assesses the relationship between household characteristics and the household migration indicator contained within the dataset. CHAID analysis is possible by using the AnswerTree extension in SPSS and the rationale for using this type of hierarchical analysis is the same as when binary logistic regression was used in the process to determine the constraint variables for the microsimulation; the variables contained within the household SAR are categorical (HOSMER, D. W. and Lemeshow, S., 2000).

The decision tree tool used is able to create a manageable set of clusters detailing the likelihood of movement for each type of household. The clusters are organised as nodes on a hierarchical tree, each variable breakdown representing an additional characteristic of the cluster; an additional branch. Each branch segment produced in the decision tree is mutually exclusive and represents a set of conditional probabilities (MAGIDSON, J., 1993); that is, if family type = 'Lone parent' then the propensity to move = 5% (0.05) per annum. Such information is based on the 2001 census variable which is derived from the question about a household's movement behaviour in the last year. Using the '*household migration indicator*' as the dependent variable and the remaining variables as the independent variables, the CHAID analysis can begin; the decision tree is allowed to grow to its maximum depth and requires at least 100 cases (household records) to be available for analysis at the parent node (node 0) and 50 cases (household records) at each successive child node. Statistical significance is set to 0.05 and the variables with the highest chi-square values are added to the tree. In this way, the tree is simplified showing only the most significant variables affecting the decision to move. Note that all variables in the household SAR are used in this analysis, only significant variables appear in the results. The resultant decision tree is shown in **Figure 4.4** below:

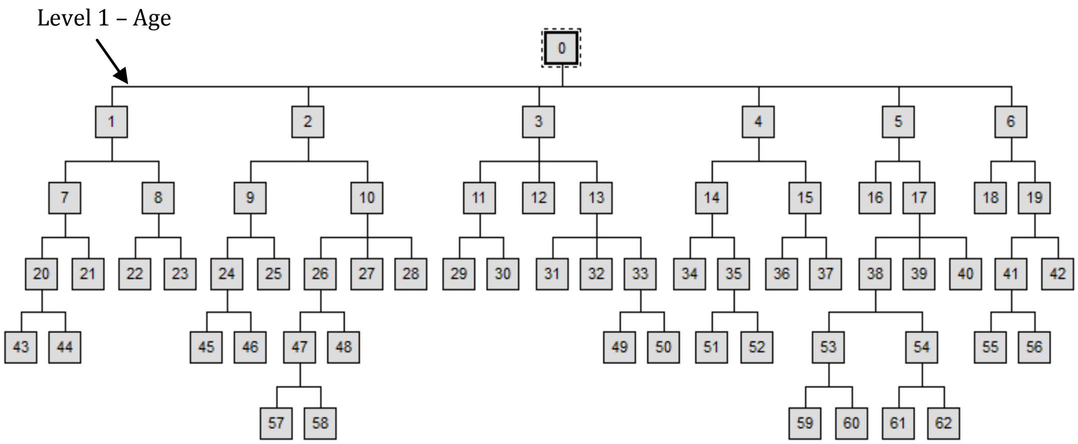


Figure 4.4 Full Decision Tree

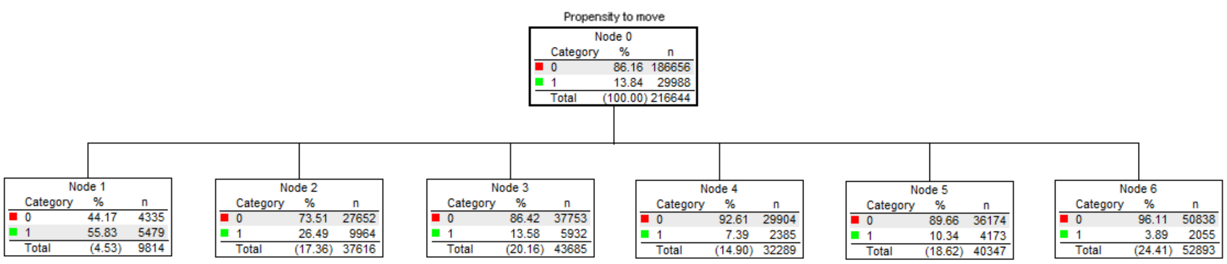


Figure 4.5 Node 0 of decision tree further defined

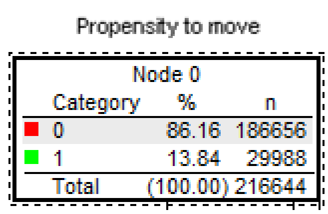


Figure 4.6 Level 1 of the Decision Tree showing Age Category

In the decision tree above, there are six primary branches on the tree. Examining the results at Node 0 (**Figure 4.5**), the statistics show that for every 100 households analysed, ~14 (13.84%) have moved. Moving down to level 1 of the decision tree, age is identified as the factor which most influences the decision to move. The use of age in the decision tree is significant. Recall that this variable was ruled out of the constraint analysis used to generate the synthetic population; its use here is thought to compensate for the earlier exclusion. Node 1 (**Figure 4.6**) represents the age category '16-25' which have an overall probability of 55.83%. This statistic gradually becomes smaller as the age of households increase. For example, at node 6 where the age category is greater than 65, the propensity to migrate is 3.89%. This observation is commensurate with the notion that younger people have a higher probability of moving than older people.

As the decision tree is extended, more levels are revealed. For each child node shown in the figure above, the tree is further subdivided to show the likely factors that influence the decision to move for each age category. This forms branches of the decision tree. Each branch, from the top of the tree to the bottom can be recognised as a segment. These segments define different types of households and detail the likelihood that different types of households will move. Though each household can only occur in one complete cluster branch, variables can be repeated from cluster to cluster. For example, the tenure variable may appear in both the '26-35' and '46-55' age categories though household X would only appear in one of these groups.

The tree is extended to reveal that for Node 1 (age 26-35), tenure is the next most important factor when the decision to move is considered. Of the 9814 households aged 26-35, 5550 live in owner occupied housing or council owned housing and 2433 households chose to move in the last year. High movement is detected in the 'Private rented or lives rent free' tenure type, where 71% (3046 households) chose to move. Furthermore, for those living in 'Private rented or lives rent free' tenure, the next most important deciding factor is the family type followed by the number of residents in the house.

Subsequent branches of this decision tree can be interpreted in much the same way. Every parent node is noted to be a significant factor with the nodes closer to the top of the tree having higher chi-square statistics than the proceeding nodes. Overall, the decision tree presented in **Figure 4.4** illustrates that for this dataset, age is the most important factor for consideration when households choose to move. This supports the notions by Boehm (1982) and Dieleman (2001) who stated that the family life cycle stage influences the decision to move and this life cycle is strongly related to the age of the individuals. Also, as emphasised in seminal work of Rossi (1955), this decision tress structure is strongly linked to the family life cycle. With age being the most influential determinant of the decision to move and other characteristics such as the family type, housing tenure, social class, house size and the number of rooms required these determinants are echoed in the work presented in **Section 2.2.1** where life course triggers and discussed.

A full description of the decision tree is presented in the **Appendix B**. In the model, the decision tree is traversed for each household to determine the likelihood of movement. If the household decides to move, a new house must be found.

4.3.3.2 Finding a new home for household *X*

Finding a new home is dependent on the behavioural rules discussed in **Section 2.2**. Recall the seven rules which were chosen based on the literature and repeated in **Table 4.7** below.

Rules	**Definitions**
Known areas	Households moves to area to which they are familiar
Index of Multiple Deprivation	Households move to better quality areas
Number of rooms requested	Households move to houses where the number of room is satisfactory
Ethnicity preference	Households move to areas where the ethnic make-up is tolerable
Transport routes available	Households move to areas where transport routes are readily available.
Socioeconomic status	Households move to houses they can afford.
Schools in proximity	Households containing school-aged children move to areas where schools are accessible.

Table 4.7 Behavioural rules used to define household location choice

Noting that in reality when trying to find a new housing unit households search a list of vacant houses amassed from various sourcing such as letting agencies, a list of 50 houses are first chosen at random to simulate this process (**Step 4bi**). An alternative to this would be to rank the list of all vacant houses by order of most suitable for each household wishing to move. This proved to be computationally expensive, however, extending the execution time of the model to over 15 hours for one simulation run. Once chosen, these 50 houses are examined based on their attractiveness to the household. Attractiveness is determined by traversing each of the seven behavioural rules.

Measuring the Attractiveness of a house

Programmatically, a function is created to execute the behavioural logic for each of the seven rules. Each function returns a value ranging from 0 to 1 where 1 indicates that the house under observation is very attractive while 0 indicates that the house under observation is not very attractive. The attractiveness value is a result of the conditions within each function. It is defined generally as follows:

$$Attractiveness = 1 - \left(\frac{|x - y|}{y}\right)$$

Where x and y represent the observed and desired values respectively; ranging from number of rooms to distance measures. In this way a house with the required characteristics is thought to be more attractive than the house which does not meet the required characteristics such as tenure type and room required. Each function returns an attractiveness value. The house with the highest summed attractiveness value is the house that is selected out of the 50 houses selected for observation. In this way, every household desirous of moving will be relocated, though some moves may prove to be more optimal than others. This is thought to mirror a similar activity in the real world.

Finding a new house using the Behavioural Rules (**Step 4bii**)

The seven behavioural rules are defined and initial examples of correct execution are illustrated. The latter process is often referred to as verification. Verification ensures that the model is working as defined (NIAZI, M. and Hussain, A., 2009). For example, if households are intolerant to ethnic diversity, this is likely to result in segregated communities. In a similar way, households with dependent children are likely to cluster within closer proximities to school than households without dependent children.

The model is run using the full EASEL area. Each rule is defined algorithmically then practical examples of correct execution are given. **Household 12493** is observed for each behavioural rule. The household is led by a 40 year old individual, occupying social housing. Living in a purpose-built flat, this White British household has no dependents and works in a lower managerial occupation. The mobility behaviour of this household is traced over a 60 year time period; initial results are presented using this time frame. Though this time frame is used, it should be reiterated that is no ageing of the population in this model and household 12493 is representative of a household with the described attributes over the 60 year period. Household 12493 will at times be referred to as the White British household.

Check Known Areas

The known areas rule simulates a household's knowledge of the community. More practically, household A has knowledge of neighbourhood B if the household once occupied a house in neighbourhood B. Using a 2 mile buffer to define a neighbourhood, if the newly found house falls within a known neighbourhood then the house is thought to be attractive. Note that **Section 2.2.2** highlighted a buffered distance of 5 miles. The EASEL area is approximately 2 miles in diameter, since the model does not include districts outside of the EASEL area, a 5 mile buffer would not capture more houses than the 2 mile buffer. For this reason the 2 mile buffer is used. Also, the buffered distance measure adopted for this rule is purely based on the literature discussed in **Section 2.2.2** as opposed to selecting specific surrounding OAs or using other geodemographic techniques to determine the areas that are familiar to households. One example of such a technique is the Output Area Classification by Vickers and Rees (2007) where output areas with similar socioeconomic characteristics are classified in the same groups. The rule is defined below using the distance measure based on the work of Rabe and Taylor (2010):

```
Check the array list storing previously occupied houses of
current household

Calculate distance between each house in list and the new
house

if (distanceApart <= 2 miles)

      attractiveness = 1

else
```

```
        attractiveness = 1 - ((Math.abs(distanceApart – 2
        miles))/ 2 miles)
    end if
```

In this algorithm, if the new house is within the buffer distance then this house would be the best house to choose, however, if the house is not within the 2 mile buffer distance then the house closest to the buffer distance will be selected. Using these conditions, the function returns the best house.

In **Figure 4.7** below, the mobility behaviour of the White British is observed. Using a 2 mile buffer around each house, the figure illustrates that each of the new houses was located within the buffered area defined by the rule. **Table 4.8** below shows the movement of White British household and the distance away from the previous house. The first three household Ids are colour coded to match the point objects in **Figure 4.7** which illustrates the first three moves using buffered distances. All distances are noted to be within the 2 mile distance from the previous house as defined in the algorithm illustrating how the rule works in isolation to all other rules.

It should be noted that the actual EASEL district is approximately 2 miles in diameter meaning that under the Known Areas rule, households that have moved from one place of residence to another are assumed to be familiar with a large percentage of the entire district at any given time.

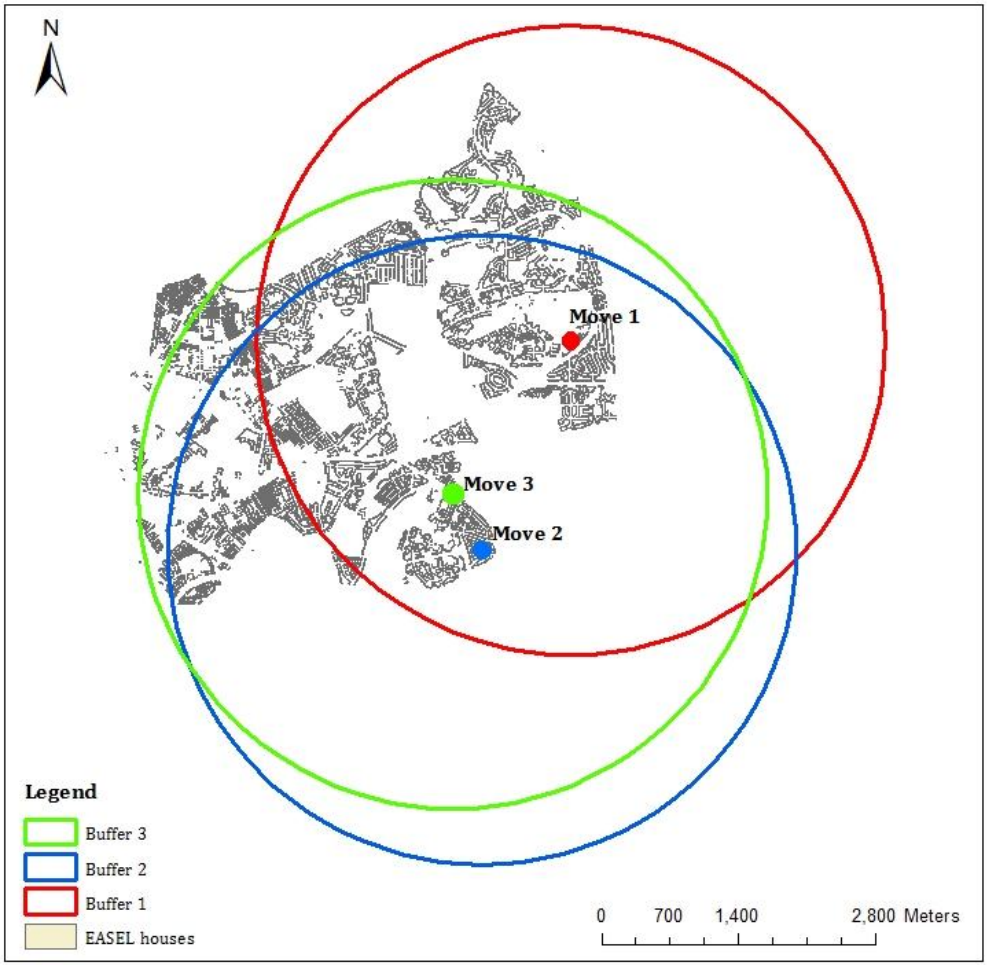


Figure 4.7 Observing the movement behaviour of the White British household based on the Known Area rule. Associated buffers shown for the first 3 moves

Household ID	Order of moves	House ID	Distance from Previous house (miles)
12493	**1**	**14550**	
	2	**19579**	1.45
	3	**18622**	0.4
	4	20704	1.14
	5	23631	1.46
	6	6623	1.82
	7	23438	1.47
	8	22727	0.82

Table 4.8 Order of moves for household 12493 under the Known Areas rule

Check Neighbourhood Quality

Based on the notion that households generally move to improvement, this rule assesses whether the new house is located in an OA that is better than the OA of the previous house. OAs are compared based on the IMD. The IMD is a formal measure of deprivation used to assess neighbourhood quality. It encompasses neighbourhood characteristics based on housing quality, crime, income and other features (LEEDS CITY COUNCIL, 2005). Note that IMD statistics for 2004 have been used. If the IMD of the new house is smaller than the IMD of the currently occupied house then it is taken to mean that the area is better. This is defined in the algorithm below.

```
if (IMD of new house < IMD of current house)
      attractiveness = 1 – ((absolute value(IMD of new house –
      IMD of current house))/IMD current house);
else attractiveness = ((absolute value (IMD of new house – IMD
of current house))/IMD of current house)*0.04;
end if

if (attractiveness < 0) attractiveness = 0;
```

Table 4.9 below illustrates the movement pattern of the households under observation. Recall that each household attempts to find the best housing alternative out of a list of 50 houses. Thus as noted in the table, there is a gradual improvement in the type of area chosen to live in until an area of comparably high IMD (77.05) is selected, despite this, the decline in IMD is once again observed after this. Though this interferes with the goal of moving to improvement, due to the closed nature of the model; static household structures and IMDs, in order for some households to move to improvement other households are not able to do this. Theoretically this is not far from reality, that is, though a better residence may be the ultimate goal, not all households move to better living conditions due to various constraints.

Household ID	House ID	Output Area	IMD
12493	14550	00DAGE0019	71.56
	18476	00DAFF0016	32.63
	9065	00DAGB0018	30.26
	24525	00DAFM0062	19.27
	52124	00DAGE0021	77.05
	12879	00DAGE0003	71.56
	52932	00DAFF0054	70.33
	20025	00DAGB0006	69.68
	68188	00DAFF0060	67.35

Table 4.9 Movement Pattern of the White British household (Neighbourhood Quality rule)

Check Room Requirements

Here the rooms required variable contained within each household record is used to determine if the newly found house has an acceptable number of rooms out of the list of 50 houses selected. Algorithmically, this rule is defined as follows:

```
if (number of rooms in new house = rooms required)
      attractiveness = 1;
else if (number of rooms in new house < rooms required)
      attractiveness = 1 - ((Math.abs(number of rooms in new
      house -
                        rooms required))/rooms required);
else attractiveness = 1 - (((Math.abs(number of rooms in new
house -
                        rooms required))/rooms required)*0.4);

end if
```

If a house is found with the same number of rooms required then this is the best house. However, if a smaller or bigger house is found, the house with the number of rooms closest to the number of rooms required will be selected as the best house. Note that if a larger house is found, the function used to calculate attractiveness will produce values outside of the 0 to 1 range, therefore this value is multiplied by 0.4. In this way, the function produces values within the 0 to 1 range.

Table 4.10 illustrates the movement behaviour of the White British household. This household requires 3 rooms; the rooms required variable is part of the attribute data associated with each household derived from the household SAR. The moves show that most of the houses acquired by the household contained 3 rooms and in some instances 4 rooms. Though the latter option was above the requirement for this household, using the algorithm, out of the 50 houses observed, the house with a room count closest to that required was selected.

Household ID	Rooms Required	Rooms Acquired	House ID
12493	**3**	3	14550
		4	13642
		4	23445
		3	5584
		3	22614

Table 4.10 Movement pattern of the White British household (Rooms rule)

Check Ethnic Mix

The ethnicity rule attempts to relocate households to an area where there are a high percentage of households of similar ethnicity. During the execution of the CHAIRS simulation, ethnicity proportions are generated by OA at the beginning of each year. The general assumption used here is that changes in the population are more noticeable over significant time periods rather than shorter time periods. In this case, calculating proportions on a yearly basis is thought to be more useful than monthly calculations. Proportions are calculated for general ethnicity groups, White, Asian, Black and Other ethnic group. In general, less that 8% of the EASEL population can be classified as non-White. The groups are detailed in the table below as follows (**Table 4.11**):

Generalised Ethnic Group	Original Subgroups	Counts of Households	Proportion of Total Population (%)
White	White British, White Irish, White Other	32908	92.1
Asian	Indian, Pakistani, Bangladeshi, Other Asian, Mixed White and Asian	1132	3.17
Black	Black Caribbean, Black African, Other Black, Mixed Black and White Caribbean, Mixed Black and White African	1335	3.74
Other	Chinese, Other Mixed, Other ethnic group	354	0.99

Table 4.11 Ethnicity groups sub-categorisation

The ethnicity rule is executed using the algorithm below. A house is most attractive if the proportion of households in its OA is greater than or equal to some preference level when compared to the proportion of households of the generalised ethnicity group of the household under observation. If the proportion is lower than preferred, the highest calculated proportion is used to select a new house. The preference level initially used is 33%, an adaptation of the Schelling model.

```
Traverse through list of all output areas
Calculate the proportion of households in output area of new
house of the same ethnicity as current household
if (proportion >= preferenceLevel) weight = 1
else weight = 1 - ((Math.abs(proportion
            preferenceLevel))/preferenceLevel);

end if
```

Table 4.12 notes the mobility behaviour of the White British household. Here, the household moves to areas where there is a high concentration of White households. The concentrations are noted to be over 80% for all cases of this example. This table can be contrasted with **Table 4.13** where **Household 12543** is observed. This household of Black Caribbean descent moves to areas where the concentration of similar households is below the preference threshold of 33% as dictated by the rule because of the lack of suitable areas. It should be noted that this preference threshold is further explored in the context of optimisation, **Section 5.3.3**. **House 12543** will be referred to as the Black Caribbean household hereafter.

Household ID	Ethnicity Subgroup	Output Area	Generalised Group Proportion (%)
12493	White British	00DAGE0028	n/a
		00DAGE0045	89
		00DAFM0046	95
		00DAGF0050	91
		00DAGB0046	93
		00DAGF0074	89
		00DAGB0010	93

Table 4.12 Movement pattern of the White British household (Ethnicity rule)

Household ID	Ethnicity Subgroup	Output Area	Generalised Group Proportion (%)
12543	Black Caribbean	00DAGF0006	n/a
		00DAFF0012	1.4
		00DAGB0058	3.73
		00DAGB0022	3.6
		00DAFM0006	8.1

Table 4.13 Movement pattern of the Black Caribbean household (Ethnicity rule)

Check Transport Routes

This rule is especially important for those households without cars. It assumes that for a household without a vehicle, a major road must be located within 1 mile of the new house. If the new house is at least 1 mile away from a major road then the new house is thought to be more attractive than a house farther away. The rule is defined as follows:

```
if (number of cars = 0){
      Traverse through the list of roads
      Calculate the distance between new house and all major
      roads
      if (distanceApart <= 1 mile)
            attractiveness = 1
      else attractiveness = 1 - ((Math.abs(distanceApart-
                        distance))/distance);
      end if
end if
```

Using this rule, only households without cars will be processed by this algorithm. A new house found within 1 mile of a major road is regarded as suitable and therefore is assigned the highest attractiveness value. In lieu of this, the house closest to a major road is selected; this may be further than 1 mile in distance. Households with cars are automatically ignore in this rule, instead they are randomly assigned a house anywhere in the EASEL district as it is assumed that with their own transport, the importance of the transport route rule is less important.

Table 4.14 shows the list of houses which the White British household subsequently occupied. This household does not own a vehicle. **Figure 4.8** shows the actual moves of this household. Notice where the major road is located; highlighted in red, and the 1 mile buffers drawn around this major road. This non-standard buffer is derived by creating 1 mile buffers around each point on the line feature representing the road. Note that in shapefile analysis, a line is a compilation of point features. The rule suggests that households without cars will try to find houses within the buffered areas. Both the table and figure shows the progressive move of the White British household which starts at house 14550 and gradually moves closer and closer to the buffered area.

Household ID	Location	House ID
12493	**1**	14550
	2	23513
	3	11624
	4	5155
	5	5187
	6	864
	7	3164
	8	4451

Table 4.14 Movement pattern of the White British household (Transport rule)

Notice that the houses in location 7 and 8 (**Table 4.14**) are the only houses within the buffer zones. Prior to these selections, other houses outside the buffer zones were selected, however, recall that houses are selected based on a subset list of 50 randomly selected houses. Though these houses do not fall within the buffer zone, the houses selected are the available houses closest to the buffer zones.

Figure 4.8 Movement pattern of White British households using the transport route rule. Buffer zone also illustrated

Check Housing Tenure

This rule simulates upward mobility on the housing ladder. An owner is more likely to search for another house which can be bought and least likely to become a social housing tenant though they may opt to go on the private rental market. In a similar way, a private renter is very likely to either continue on the private market or purchase a home, though social housing may be another option. Finally, a social housing tenant is more likely to continue in social housing, and less likely to purchase a home, though such a tenant may opt to go on the private renting market.

Thus the attractiveness values are assigned as follows, where tenure 1, 2 and 3 represent ownership, social rental and private rental respectively:

```
if (currentTenure = 1)
```

```
        if (newTenure = 1) attractiveness = 1
        else if (newTenure = 2) attractiveness = 0.33
        else if (newTenure = 3) attractiveness = 0.66
    else if (currentTenure() = 3){
        if (newTenure = 1) attractiveness = 0.5
        else if (newTenure = 2) attractiveness = 0.5
        else if (newTenure = 3) attractiveness = 1
    else
        if (newTenure = 1) attractiveness = 0.33
        else if (newTenure = 2) attractiveness = 1
        else if (newTenure = 3) attractiveness = 0.66
    end if
```

Notice the weights assigned to each condition follow the qualitative likelihoods earlier mentioned. Also note that the actual numerical values chosen have been arbitrarily chosen though they follow the same likelihood trends outlined. Thus owners are most likely to resume ownership. This trend follows through for private and public renters who are most likely to resume private and public rentals respectively. The attractiveness varies accordingly if they are faced with another tenure option. It is important to realise that it is through this rule that the social housing element in the area is modelled. Current government policy leans heavily towards moving households from the social housing 'market' into private or intermediate housing (**Section 2.3**). As such, this rule replicates this driver by including the option for people to shift from social housing tenancy to the private sector. However, the decision-making dynamics of those moving within the social housing or intermediate sector are more complicated (for example, they often involve movement between distant social housing areas), and they are not represented within this model – it is assumed that for social housing, the area is neither a net source or sink of people to or from other social housing areas in the city. This is generally true for the case study area used here, but would need adaptation for other regeneration areas.

The White British household reside in social housing. **Table 4.15** shows the mobility behaviour of this household. Here the household continues as a social housing tenant only moving to private accommodation once through the 60 year model run.

Household ID	Old Tenure	New Tenure	House ID
12493	**2**	2	14550
		2	53060
		2	7382
		2	54839
		3	56190
		2	13158
		2	16385

Table 4.15 Mobility behaviour of the White British household (Tenure rule)

Check Schools in Proximity

This rule assesses whether there is a school within close proximity to the new house. A 3 mile distance radius is used to determine how attractive each vacant house would be. This distance is based on the statutory walking distance rule reported by the government (SCHOOL ACCESS SERVICES, 2011). Using the statutory walking distance, children under the age of 8 years old, living more than 2 miles away from their school qualify for free transport to and from school. The same applies to children over the age of 8 years old living more than 3 miles away from their school. This suggests that in general, distances between 0 and 3 miles of a school are preferred.

Note that, since data is used at the household level, details on the number of dependents is the only data recorded; there is no data on the age of dependents. For this reason, this rule makes no distinction between primary and secondary schools, if there is a school within a 3 mile radius of the new house then this house is thought to be more attractive than a house not within such a distance. The details of the rule are defined below:

```
if (numberOfDependents > 0){
      Traverse through list of all schools
      Calculate distance between school and house

      if (distanceApart <= 3 miles)
            attractiveness = 1
      else
            attractiveness = 1 - ((Math.abs(distanceApart-
                                  distance))/distance)

end if
```

The rule is only traversed for households with dependent children. If the new house is within 3 miles of a school then this house is most attractive, however, the new house closest to a school will be selected. The Black Caribbean household has 2 dependent children. **Figure 4.9** shows the distribution of schools throughout the EASEL district. The schools have been highlighted using red point objects. Note that the algorithm does not distinguish between school qualities. While this is not problematic in the EASEL area, this factor would need further refinement for more general simulations.

There are nine schools reasonably spread across the EASEL area, considering that the entire district is less than 3 miles in diameter all households live within this distance of a school. This is also illustrated in **Figure 4.9** where a 3 mile buffer is drawn around the school object coloured purple in the middle of the EASEL district.

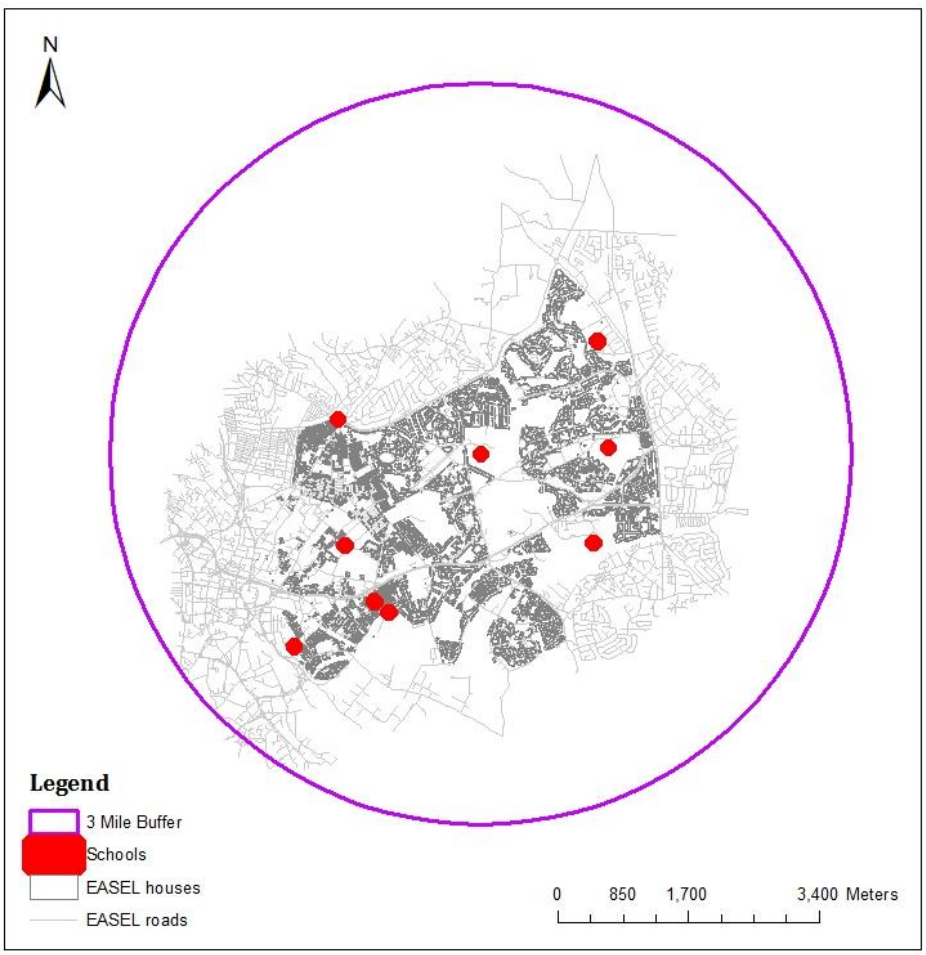


Figure 4.9 Illustrating the 3 mile buffer around the EASEL district

Table 4.16 shows the mobility behaviour of the household when this rule is applied. Here the household is noted to move to houses which are less than 3 miles within the location of schools.

Household ID	House ID	Distance from School (miles)
12543	15957	0.7
	2505	0.9
	11358	0.5
	16251	0.3
	24031	0.54
	718	0.24
	8258	0.2

Table 4.16 Movement pattern for the Black Caribbean household (School rule)

The rules illustrate the behaviour of the household agents interacting with the environment in which they exist. This environment includes the OAs, roads, schools and houses. On observation of the rules and results, it may be expected to see some measure of improvement on household circumstance in each behavioural rule. For example, using the ethnicity rule, it may be thought that a household should only move to areas which are more concentrated than the previous neighbourhood in which the household lived. Though this may be the case, it should be reiterated that households choose from a list of 50 houses each time a residential mobility decision is made, thus each household makes the best possible decision with the information available. It is also assumed that each household has to move once selected in the CHAIRS algorithm.

4.4 Model Outputs

Though Repast provides a pictorial display for each model run, due to the large volume of data processed in the CHAIRS model this pictorial display could not be used during the execution of the full model. For this reason, when the model is executed data is written to output text files. This data is then used to analyse the behaviour of the synthetic population. Household locations are the main output generated. This is printed on a yearly basis until the simulation ends. Counts of the population by OA are compared to counts of the population by OA in a carefully selected validation dataset.

Data is also georeferenced so as to represent the population distribution using maps. Model outputs are discussed extensively in **Chapters 5** and **6** where calibration/validation and the final results are discussed respectively.

4.5 Model Assumptions

The CHAIRS model has been created on the basis of a few assumptions. Some assumptions were needed due to a lack of more detailed information, other assumptions were made so as to make the model processing and output more manageable. There are eleven assumptions, as noted below.

1. The CHAIRS model is represented as a closed system; there is no migration beyond the boundaries of the EASEL district. This means that the model allows for the reshuffling of EASEL households over the model execution period. Also, households are unlikely to move far distances as those with jobs are more likely to work in the city centre which is the neighbourhood district of EASEL.

2. Based on the household SAR statistics, 14% of the population migrated in the year preceding the census (2000-2001). This statistic is generated based on the census question which asked households to state whether they have moved house over the last year. Thus for each simulated year in the CHAIRS model, 14% of the population is allowed to move.

3. Households are the individual level data used within the model. The alternative would be to represent each person in the EASEL district then aggregate these individuals in to households. However, the assumption has been made that when a household wishes to move then the entire household is affected therefore it is sufficient to represent the household as one entity without having to represent individuals in the household.

4. There is no mechanism for increasing or reducing the EASEL population through births and deaths. It was felt that building a mortality/fertility model would be a large undertaking and distract from the main focus of the CHAIRS model, that is, to build a general model of residential mobility. Having said this, though there are no births and deaths, the propensity to move statistic encapsulates mortality/fertility behaviour by representing the probabilities of households of differing ages to migrate.

5. The seven behavioural rules used are sufficient to recreate a general model of residential mobility. Residential mobility behaviour for the EASEL area is no different to general residential mobility behaviour for those living in any other area of the country. All households are faced with the same types of decisions, such as affordability and distance to work etc. It is the household's circumstance that varies – this is where the decision for one house over another may vary. For example, a household living under ownership is more likely to move to another situation of ownership than to the social housing market though this is not impossible. These behaviours are recreated in the behavioural rules used in the CHAIRS simulation.

6. The propensity to move statistic is sufficient to determine the probability that a household would decide to move house. It encapsulates the behaviour linked to the family life cycle. In this way, younger persons are noted to have higher propensities to move than older persons. As a consequence, the types of households that move follow this same trend in the CHAIRS model.

7. Mechanisms that drive the housing market such as mortgage interest rates and house prices are not accounted for in the model, rather there is a larger focus on the residential mobility behaviour.

8. A list of 50 vacant houses is chosen from the list of all vacant houses. This is used to represent a process of households finding out about vacant houses via various types of media. Only by testing the attractiveness of each of the vacant houses can the household determine the most suitable house.

9. Once selected by the CHAIRS algorithm, the household must move. This ensures that 14% of the population can be counted and used to represent a 1 year time period in the model. Thus the attractiveness functions used select a new house ensures that of the 50 vacant houses selected, the most attractive is chosen. The alternative would be to indicate that none of the selected homes is attractive. In reality, this may be representative of a process of compromise which some households go through.

10. Vacancy chains are not in operation therefore meaning that once a household leaves a house, it can be reoccupied immediately. Thus there is no need for this additional process to be modelled as part of the CHAIRS simulation.

11. The IMD is used as an indicator of neighbourhood quality as it encapsulates several neighbourhood qualities namely barriers to housing, education, crime, income and other features. This complex indicator is thought to be sufficient in describing the quality of each OA.

4.6 Conclusion

In this chapter the rudiments of the CHAIRS simulation have been discussed in the context of inputs, processing and outputs of the model. Much is discussed on the derivation of the individual level dataset used to represent households. This dataset was created through a process of microsimulation using original data from the 2001 census and the 2001 household SAR. The shapefiles used to represent other entities used in the model namely OAs, schools, roads and houses were also discussed. The overall details of the processing phase of the model were also presented. Each behavioural rule used to govern residential mobility behaviour is presented and results are used to verify the correct working of each rule. General model outputs are then described. The chapter ends with a list of the assumptions made when creating the CHAIRS simulation.

Chapter 5
Testing the Model for Realism using Calibration and Validation Techniques

5.1 Introduction

The fundamental goal of the CHAIRS model is to assess the possible effects of some of the property-led urban regeneration projects conceptualised for the EASEL district using a model of housing choice. With the model designed and implemented, it can now be tested and optimised so that the resultant model outputs can be compared to reality. This is the deployment phase of the model development process; the stage at which the model is calibrated and validated.

Calibration is the process of ensuring that parameters in the model generate behaviours similar to those observed in the real world (DE SMITH, M. J. *et al*., 2009). Parameters may be described as any value used to adjust variables within the model. For example, distance measures are noted as parameters in the CHAIRS model as various rule-sets use the distance as an integral part of behaviour simulation. In calibrating such a model, various distance values can be systematically chosen; the model is executed several times, using a different distance value in each model run. The results from each model run are then compared to known, real-world data to assess their validity. The distance parameter which generates results that best match the results observed in the real-world is likely to be the distance parameter chosen ((LOUIE, M. A. and Carley, K., 2008); (DE SMITH, M. J. *et al*., 2009)). This latter phase of data comparison is the validation process. Validation provides proof that the model recreates reality with a satisfactory measure of accuracy ((SARGENT, T. J., 1998) in (BIANCHI, C. *et al*., 2008)).

Calibration and validation are important parts of the model development process, however, these procedures can be challenging for many reasons. Crooks *et al.* (2008) point out the difficulty of obtaining accurate individual level data due to issues of data protection and individual privacy as well as the non-existent of needed data. ABMs, such as the CHAIRS simulation, require specific variables to enable them to work. It is unfortunate that limitations on data may affect this model.

Issues such as these challenge the ability of researchers who create ABMs and, in the past, modellers may have overlooked the importance of rigorous calibration and testing as a result of this. For example, though the work of Thomas Schelling has contributed significantly to ABM literature, Schelling's original model was only loosely tested. Schelling was able to illustrate that when households exercised slight preferences related to racial tolerance, neighbourhoods became totally segregated ((SCHELLING, T. C., 1969); **Section 3.3.1**). These results showed clear parallels to the qualitative literature which suggests that household preferences influence segregation patterns across society (MASSEY, D. S. and Fischer, M. J., 2000), however, the model was not rigorously calibrated or validated to show the extent to which this theory matched reality. Despite these challenges, it is important to reiterate that simulation models are abstractions of reality (COLBURN, T. and Shute, G., 2007). Every detail in the real world, to the minutest degree, is not likely to be represented. In many cases, there may not be sufficient data to document these behaviours while in other cases, the dynamics in the real-world may not be clearly understood. The balance between realism and the simplicity of ABMs is therefore important and it is more key to ensure that there are clear parallels between the dynamics of the model and dynamics in the real world.

In recent years modellers have paid closer attention to assessing the accuracy and correctness of ABMs (BIANCHI, C. *et al*., 2008). Work by Heppenstall *et al.* (2005) and Malleson *et al.* (2008) are two examples of different but complementary techniques of calibrating and validating ABMs. Malleson *et al.* (2008) detail a process of manual calibration in a burglary simulation model while Heppenstall *et al.* (2005) used automatic calibration in a petrol pricing model. Though distinctly contrasting techniques, both have their advantages. Manual calibration allows for expert input when choosing the parameter space, while automatic calibration allows for a certain level of objectivity and can allow for a more thorough exploration of the parameter space.

In the case of Malleson *et al.* (2008), significant parameters were systematically altered and used to compare the results generated by their crime model with published statistics on burglary. In this way, trends were analysed and compared to real world trends. Malleson *et al.* (2008) note that such a technique was chosen because the complexity of the model in relation to the limitations of the available computing power did not permit for automated techniques to be used. While, the National Grid Service (NGS) supercomputing environment used allowed processes to run for up to two days, a time period well in excess of this was required by the crime model for the purpose of calibration (N Malleson 2010, pers. comm.).

Heppenstall *et al.* (2005) employed the use of a genetic algorithm to automatically calibrate a petrol pricing model. Genetic algorithms use various combinations of parameters to create the fittest or best match (HEPPENSTALL, A. *et al*., 2005). It is a process of finding the most optimal parameters and building on such parameters with the aim of finding a better solution. The algorithm employed models the biological process of evolution by treating parameters as genes. Poorly performing parameters (genes) are not used to generate new generations of parameters. However, parameters that perform well are built upon (mutated) in order to explore potentially advantageous change. The algorithm is governed by some stopping condition which is predefined by the modeller and relates to the quality with which the model predicts known scenarios. When this stopping condition is met, the solution is reported. Overall, Heppenstall *et al.* (2005) concluded that this evolutionary process was successful in predicting the long-term profitability of petrol stations.

As introduced in **Chapter 3**, other calibration/validation techniques have been used when optimising and testing models of residential mobility in an ABM context. As mentioned in **Table 3.2**, significant parameters can be toggled and tested for an acceptable goodness of fit as in the model by Benenson (2004) while validation can be enacted by comparing trends observed in the model outcomes with known trends in census data (ZHANG, J., 2004), utilising inequality indices (AGUILERA, A. and Ugalde, E., 2007), using absolute error statistics (BENENSON, I., 2004) among others. In this chapter, model calibration and validation techniques are applied to the CHAIRS simulation model with a view to optimising the model to the extent that the results of the CHAIRS simulation closely match real-world situations. Thus, a systematic approach to calibrating and validating the CHAIRS housing simulation agent-based model is first presented. This is followed by a discussion on the calibration/validation results. The chapter ends with a critique of these results outlining the major challenges in this modelling exercise. **Figure 5.1** is used to illustrate the link between OAs and LLSOAs, the latter is used when results are presented in this chapter and the chapter to follow. A full list of LLSOA codes and names can be found in **Appendix I**.

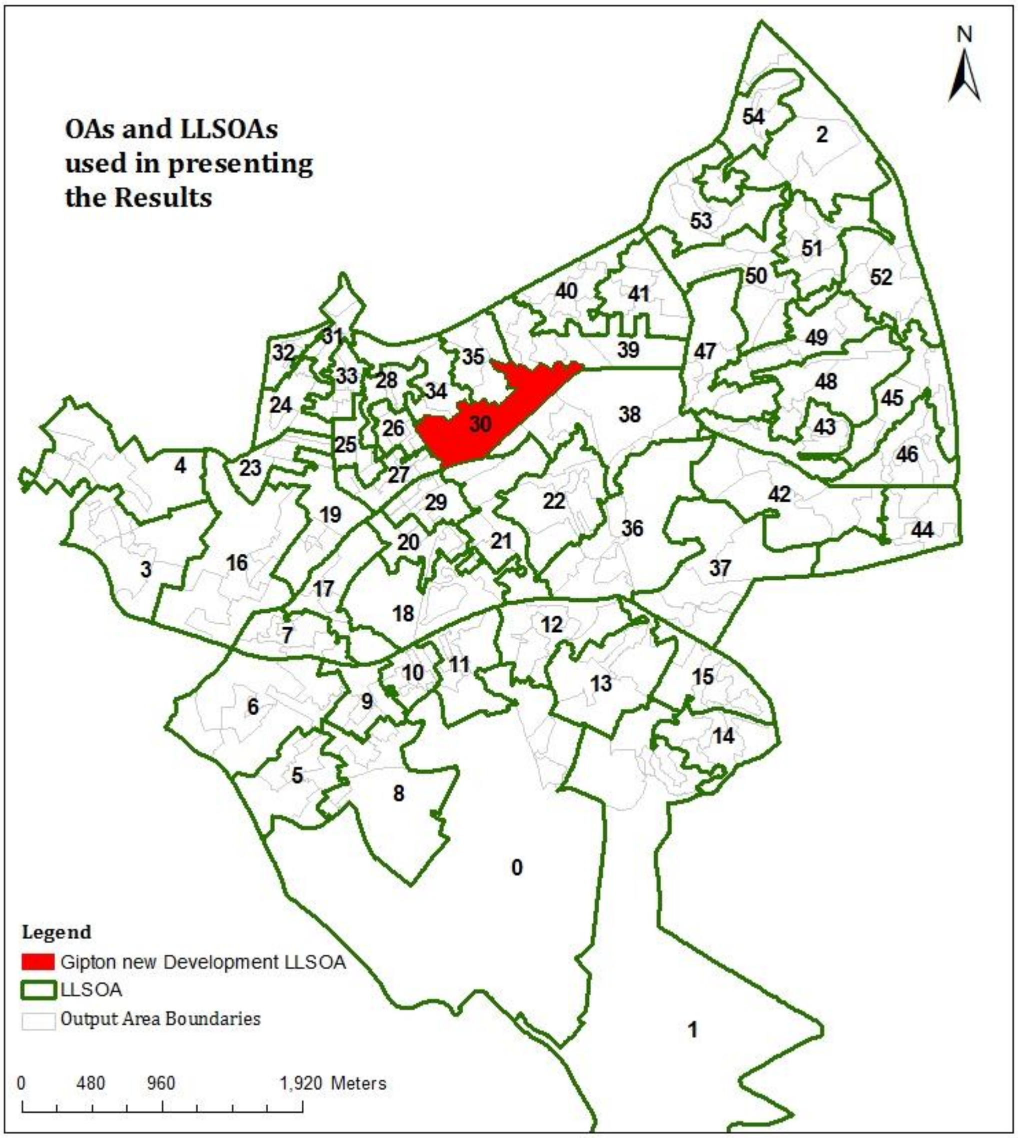


Figure 5.1 LLSOAs illustrated in the context of OAs for the EASEL district highlighting the Gipton new development area. LLSOAs labelled by generic identifier. A full list of SOA codes and titles can be found in **Appendix I**.

5.2 Defining the Methodology for Calibrating and Validating the CHAIRS Model

The CHAIRS simulation has been calibrated and validated using a systematic combination of procedures. All simulations begin as at 2001; households are distributed across the EASEL district based on the 2001 Census distribution. All simulation results are validated in the 2005/2006 period using Acxiom's ROP data for the EASEL area. When calibration/validation is completed, the simulation is executed from 2001 until 2021. The final results are presented in **Chapter 6**. All data used for the purpose of validation is detailed in **Table 5.1** below:

Dataset	Purpose
2001 Census	Census proportions used to reweight the ROP
Acxiom's ROP	Survey data compared to results from the CHAIRS simulation

Table 5.1 Behavioural rules used to define household location choice in the model

The calibration/validation procedure are defined as follows. Each step is annotated with a section number where it is defined in more detail.

Step 1 Select model parameters; identify all combinations of model parameters (rule sets); generate CHAIRS simulation results for each combination of parameters (**Section 5.2.1**).

Step 2 Identify real world data for validation, that is, Acxiom's ROP; reweight the ROP to reflect census proportions (**Section 5.2.2-5.2.7**). Since the ROP is a survey dataset, this reweighting process is important so as to make the number of records contained within the ROP comparable with the actual number of households in the EASEL area.

Step 3 Compare the CHAIRS results with the ROP results; measure the number of errors using the total absolute error (TAE) and standardised absolute error (SAE) statistics (**Section 5.2.8-5.2.9**).

Step 4 Choose the rule set with the least number of errors (**Section 5.3.1-5.3.2**).

Step 5 Optimise additional parameters contained within the chosen rule set (**Section 5.3.3**).

Step 6 Generate final results using optimised rule set (**Chapter 6**).

These steps can be illustrated using a more simple example. Imagine a model with two rules; Ethnicity and Known Areas. Based on these two rules there are three rule sets; Ethnicity, Known Areas, Ethnicity and Known Areas. Here the order of the rule sets does not matter. When the model is executed for each of these rule sets the results are compared to some validation dataset. The rule set generating the least number of errors is chosen and used in the final simulation. Imagine that the Ethnicity and Known Areas rules when used together generate the least number of errors. This is the rule set that will be chosen for the final simulation because of its ability to recreate reality better than the other rule sets executed. This simple illustration is further expanded upon in **Section 5.2.10** which presents an example of the full calibration/validation procedure using a sample of the actual model data.

5.2.1 Model Calibration: Using Parameters to Calibrate the Model

Recall from the previous chapter the seven rules used to define household location choice behaviour (**Section 4.3.3.2**). The rules, as listed in **Table 5.2** range from behaviours based on familiar neighbourhoods to behaviours based on ethnic tolerances and school proximity. For the purpose of model calibration, these seven rules will be used as binary parameters; that is, they can be switched on or off in varied combinations for each model run.

Rules	Definitions
Known areas	Households moves to area to which they are familiar
Index of Multiple Deprivation	Households move to better quality areas
Number of rooms requested	Households move to houses where the number of room is satisfactory
Ethnicity preference	Households move to areas where the ethnic make-up is tolerable
Transport routes available	Households move to areas where transport routes are readily available.
Socioeconomic status	Households move to houses they can afford.
Schools in proximity	Households containing school-aged children move to areas where schools are accessible.

Table 5.2 Behavioural rules used to define household location choice in the model

Though other parameters will also be explored, such as distance measures and the level of ethnic tolerance, using the seven rules as binary parameters allows for a diverse set of rule combinations to be created. In this way, specific rules can be omitted from some rule-sets while being included in others.

In a way this model builds on the ABM created by Schelling (SCHELLING, T. C., 1969) by adding further behaviours. Whereas the Schelling model explored the effects of ethnic tolerance, additional rules or behaviours have been added as noted previously. This modelling application is therefore not only an exercise in exploring individual housing choice preferences but also a technical exercise aiming to augment Schelling's model by testing the extent to which the model is able to create real-world results using various rule combinations. In the qualitative literature there is much to suggest that the residential mobility decision is a complex process comprising of many considerations (DIELEMAN, F., 2001), however, results from this quantitative model may suggest that the equivalent patterns of socioeconomic segregation can be produced by a model of much simpler core behaviour.

Therefore, likening the seven rules to a string of ones and zeros, where one indicates that the rule is being used and zero indicates that the rule is not being used, **Table 5.3** below gives a few examples of how the rule-sets can be combined.

Example	Transport	School	Ethnic Tolerance	Neighbourhood Quality	Familiar Neighbourhoods	Number of Rooms	Social Status
A	**1**	**1**	**1**	**1**	**1**	**1**	**1**
B	**1**	0	**1**	**1**	**1**	0	**1**
C	**1**	0	0	0	0	0	0

Table 5.3 Examples of different rule-set combinations

Example A represents a rule combination where all rules are in use. Alternatively, some rules can be switched off as in Example B where the *School* and *Rooms* rules are not used. Instead, Example B represents a rule combination where the *Transport*, *Ethnicity*, *Neighbourhood Quality*, *Familiar Neighbourhoods* and *Social Status* rules are in use. Conversely, all but one rule can be switched off as in the case of Example C where only the *Transport* rule is in use.

Notice, in these three examples (**Table 5.3**) that a certain number of rules are active each time. In Example A, seven out of seven rules are chosen to create the rule combination while in Examples B and C, five out of seven and one out of seven rules are chosen respectively. Using combination theory where the order of the rule-set does not matter and repetition of rules is not allowed, the total number of combinations can be generated from the seven rules in the following way:

Total number of combinations =

$\sum$ (number of combinations when: 1 of 7 rules chosen,
2 of 7 rules chosen, 3 of 7 rules chosen, 4 of 7 rules chosen,
5 of 7 rules chosen, 6 of 7 rules chosen, 7 of 7 rules chosen)

Here the number of combinations is calculated by using the following formulae:

$$\frac{n!}{r!\,(n-r)!}$$

Where,

n represents the total number of rules (parameters),

r represents the number of rules (parameters) chosen at a specific time, and

! is the factorial of a number.

In total, there are **127** rule combinations. These combinations will be traversed in the calibration exercise. This is practical and systematic as it simply requires one model run for each of the 127 rule-sets created. It therefore negates the need for automatic calibration as the total combination space can be traversed manually. By creating different rule-set combinations, comparisons can be made by exploring how well each rule-set is able to replicate reality, with the notion that some rule-sets may produce a better match to reality than others. A complete list of all rule combinations can be found in **Appendix C**. In general, the strength of the methodology used is that the entire combination space using the seven behavioural rules can be explored and validated. It is not necessary to resort to sampling the space to optimise and validate the model, for example using a greedy algorithm, as is done in many other simulation models (SHOOMAN, M. L., 2002).

The second part of the calibration exercise is to calibrate the additional parameters used in the model. Four of the seven behavioural rules contain additional parameters. These parameters are listed in the **Table 5.4** below.

Rules	**Additional Parameters**
Known areas	Buffered distance around current neighbourhood, e.g. 2 miles
Index of Multiple Deprivation	n/a
Number of rooms requested	n/a
Ethnicity preference	Degree of tolerance to other ethnic groups, e.g. 33%
Transport routes available	Distance from major road, e.g. 1 mile
Socioeconomic status	n/a
Schools in proximity	Distance from school, i.e. 3 miles

Table 5.4 Additional Model Parameters

Though original values were adopted from the qualitative literature, it is possible that alternative values derived from larger quantitative datasets might be appropriate. The calibration process explores this possibility and allows for further optimisation as a means of improving the performance of the chosen rule-set. For example, if the distance measure in the *Known Areas* rule is optimised, the results may show that a value of 10 miles would produce a better fit to real world data than a value of five miles as suggested by the literature. Similarly, using the *Ethnicity Preference* rule, the process of calibration may reveal that a tolerance level of 20% generates results which better fit reality as opposed to 33% as suggested by Schelling (1969). Therefore by varying the estimates as given in the qualitative literature, the extent to which reality can be recreated using alternate estimates can be examined and reflection on the literature gained. To a degree, this is a form of sensitivity testing as well, for if multiple values are appropriate for a variable without variation in other parameters, it suggests the model is insensitive to this variable. Apparently essential variables that have correlated variations may suggest missing variables or processes for which both variables are proxies. By examining variable variation it may, therefore, be possible to say something about how well a model is structured and/or how well the literature reflects the model needed.

A process like this can challenge the parameter values noted in the qualitative literature suggesting that using other values may be more useful. However, bear in mind that the model presented is an abstraction of reality and as such does not replicate every facet of the real world. It can be argued that these parameters are used to improve model performance in lieu of the behaviours not replicated. Thus, the conclusion should be that a model with, for example, the CHAIRS dynamics could use parameter x or y in order to create a likeness of reality even if this parameter does not agree with the literature.

In order to find the best parameters some mechanism to alter the parameter values systematically is needed. For example, buffered distances for the *Known Areas*, *Transport* and *Schools* rules may be set to range from 1 to 10 miles while ethnic tolerance may be set to range from 1 to 100%. In order to traverse such a parameter space, a total of 100,000 (10*10*10*100) combinations of these four additional parameters must be explored if every combination is investigated. Consider that for a thorough investigation this should be executed for each of the 127 models to provide extended evidence that the actual rule-set chosen is the best and that if each model runs for ~2 hours, a total of 12 million hours (100,000*127*2) would be needed to explore the entire solution space. This amounts to more than 2000 years! This assumes that manual calibration is used; where each variable must be manually altered to form different parameter combinations. Even if automatic calibration is facilitated by a genetic algorithm, this is still a mammoth undertaking, since a large number of parameter combinations are still likely to be explored. Consequently the necessary processing would be computationally expensive. Thus, in order to reduce this extensive amount of time, first the best performing model is chosen out of the 127 models, this model is then used to find the best combination of additional parameters through a process of manual calibration.

Another consideration when optimising the model is the identification of relevant rule-sets which match the characteristics of each household represented. What is important to note is that though it is possible that each of the 35000 households represented in the EASEL district may behave differently, it is not probable. Households may all have different characteristics as defined by the individual-level input data (SAR) but as suggested by the literature, they exhibit the same kinds of behaviours when making location choice decisions. What makes the difference between the decisions of one household from another, are the factors that constrain each household; for example, a household's income may limit the amount it can afford on rent or mortgage payments. Therefore, defining individual rule-sets for each household is not necessary. For this reason, rule-sets for different types of households are also not implemented. Likewise, rule-sets can be created for various household types though not implemented in the model, though this may be an oversimplification of the situation; different ethnic groups may have different views on long-term debt arrangements and therefore influence their choice for ownership via mortgage agreements or the rental market (GJESSING, M., 2010).

In this way, household constraints are used to differentiate the behaviour of one household from another. For example, a council tenant desirous of moving is likely to be limited to living in council housing as opposed to a household living in private market housing choosing to move from rented accommodation to ownership. Such constraints are managed in the rule definitions (**Section 4.3.3.2**).

Once all of the combinations of the seven rules (parameters) are determined, each combination is executed in the supercomputing environment. Each model may run for approximately one hour in order to generate results starting from 2001 and running to 2006, though the execution time is dependent on the complexity of the rule-set activated. For example, a rule set containing the transport rule where distance calculations are to be made is likely to take longer than a rule-set where no distance calculations are made. A complete list of household attributes is outputted from the model and this list is used in the validation process.

When the behaviours are optimised and validated to confirm the degree to which they match reality, the resultant model can then be used to run scenarios as in the chapter to follow; **Chapter 6**. It is the results of the scenarios that are most important in assessing the outcomes of the EASEL regeneration scheme, the goal of this project.

Thus far the calibration process has been explained. First, all possible rule combinations of the seven behavioural rules are identified. The model is executed once for each rule combination and the results validated. Once the best performing model is identified, a similar process if repeated for the additional parameters in the model such as distance measures. Again the results are validated in order to identify the best performing parameter combination. The validation compares the results of the model with real world data. This process is now discussed.

5.2.2 Model Validation: Choosing a Real-World Dataset for Validating Model Results

In the absence of a national census post 2001, there are a number of datasets which can be obtained for the purpose of validating this model. Some of these include the British Household Panel Survey (BHPS), the General Household Survey (GHS), the English Housing Survey (EHS), the ROP and the Pupil Level Annual School Census (PLASC). Barring the PLASC, each of these datasets contain, with varying degrees, the necessary demographic details as well as details on housing types and residential mobility behaviour. All of these datasets are collected at regular time intervals, dating as far back as 1971 in the case of the GHS.

Used largely for research purposes in the public sector domain, the BHPS, GHS, EHS and PLASC appear popular, having been used to inform studies ranging from Poverty and Exclusion to Labour Market Mobility and Neighbourhood Change ((DOWNWARD, P., 2007); (CONTOYANNIS, P. *et al*., 2004); (BUTLER, T. *et al*., 2007)). Survey sample sizes range from 5500 respondents for the BHPS, 13000 respondents for the GHS, to approximately 17000 respondents for the EHS. Many of these surveys, though administered in specific regions of the UK, are used to represent behaviour across Great Britain. The PLASC dataset is slightly different in composition as it captures demographic details on individual school-aged children within every Local Authority. It does not, however, contain any household specific data. For this reason, this dataset is limiting as details on household characteristics are needed in this optimisation exercise.

One dataset that is less used and new to the academic domain is the ROP provided by the company Acxiom. Originally called the National Shoppers' Survey, the ROP is a privately collected dataset reporting on household consumption and expenditure. The survey provides a dataset of household data such as demographic details, spending habits and lifestyle choices. The data is used to enable the company's mission of transforming data collected from different sources into actionable information which can help clients understand customers preferences, improve customer acquisition and retention, predict consumer behaviour and locate optimum retail sites (BłASZCZYńSKI, J. *et al*., 2006). The ROP is collected biannually and has been available since 2004. Containing a sample of over 1 million records for Great Britain, the ROP has been proven to compare favourably to other survey products such as the Expenditure and Food Survey (EFS), the Labour Force Survey (LFS), the BHPS, the GHS and the Survey for English Housing (SEH) (Thompson *et al*., 2010). Unlike these surveys, the added benefit of the ROP is the frequency of its availability and the inclusion of data directly from EASEL area residents. In this respect, it is a valuable source of information in lieu of a more recent UK Census. In short, the ROP presents a fresh alternative to the more commonly used datasets in the academic domain and is favoured for the validation.

5.2.3 Describing the Research Opinion Poll (ROP)

Survey data for this dataset is collected biannually in the months of September and January. These two data collection points are used to represent a full year of data. For example, data collected in September 2004 is concatenated with data in January 2005 to create a dataset representative of 2005 data (THOMPSON, C. *et al*., 2010). Acxiom then uses this data to create aggregate products such as the Acxiom Population Estimates, Aggregate Data and its Behavioural Geodemographic Classification System called PersonicX. The methodology for the creation of each of these aggregated datasets is preserved under confidential cover. What is available for use in this optimisation exercise is the original raw survey data, otherwise known as micro data.

According to Thompson *et al.* (2010), the ROP contains biases. Comparing the ROP with the 2001 UK Census would reveal that there is a generally low response rate from London Centre, Cosmopolitan and Suburban areas. In contrast, there appears to be a high response rate for council renters and affluent elderly people in rural communities. Consistent with these biases is the low response rate among better-off executives in inner city areas and white collar workers in better-off multiethnic areas. Conversely, the response rate among individuals of mining and manufacturing backgrounds is high. One may argue that those living in the busy London Centre are likely to be younger individuals who are less likely to have the time or the requisite interest needed when completing such a survey.

With regards to actual demographics, more females than males complete the survey. Added to this, there is an over-representation of respondents aged 40-70 years. This highlights the under-representation of respondents on either side of this boundary (< 40 years, > 70 years) with the under-representation of the 18-24 age groups noted to be higher than any other age group. Though a good response rate is recorded for Yorkshire and the Humber, there is an over-representation of White British households and a low level of BME groups; namely Asian and Black households. This disparity in the representation of minority groups is not uncommon in surveys, however. Gibson *et al.* (1999) argue that ethnic minority groups are largely skeptical and much harder to engage due to the language barriers and the fact that they tend to be more marginalised than mainstream society.

These biases are both advantageous and disadvantageous when comparisons are made to an area such as EASEL as they overlap with the profile of many EASEL area residents. The EASEL district contains a large numbers of White British households and a significant number of council tenants; these are the groups which have high response rates. Conversely, there are also concentrations of BME groups represented in this area. Acxiom attempts to reduce these biases in various ways. Most common is the oversampling of under-represented groups, the use of online surveys to capture the attention of younger individuals and the use of tangible rewards for survey completion. In addition, though Acxiom has been successful in collecting over one million records each time the survey is distributed and its ROP dataset is rich with demographic details at high level geographies such as Wards and Cities, it appears that at lower geographies, response levels in some areas can at times be low. For example, in the EASEL area for the 2005/2006 period, there were 993 respondents recorded in the ROP in comparison to the 35000 households recorded by the census in 2001. Since specific variables are examined in this validation exercise, a cleaning process was applied to eliminate records containing null values for these. This reduced the ROP for the EASEL district to a total of 646 households.

It should be noted that other datasets such as the BHPS and GHS are not advantageous in this regard as survey samples are not collected for Leeds or EASEL, instead all data collected is used to represent the entire country. In the case of the ROP, the data used to represent the EASEL area is taken directly from that district preserving the trends in the EASEL district. A complete review of the ROP can be found in the working paper by Thompson *et al.* (2010) entitled *Understanding and Validating Acxiom's Research Opinion Poll Data*.

5.2.4 Preparing the Validation Dataset for Use: Initial Considerations

Before the ROP dataset is used, it must be altered in such a way that model results can be usefully compared to it. To improve the number of records available in the ROP micro data two procedures are adopted. First, both the 2005 and 2006 data is used for validation to increase the number of actual records. ROP data from 2005 and 2006 are concatenated to create one validation dataset. In turn, this will be compared to the concatenated 2005/2006 model results. Notice that the comparison point starts with data at year 2005 while, in fact, the model begins with data at year 2001. For the purpose of validation, the model is rolled forward from 2001 to 2005/2006 and its predictive capacity examined.

Secondly, to ensure that the population counts in the ROP are comparable to those in the CHAIRS model and vice versa, the ROP data is then systematically reweighted so it can be compared with total population predictions. For example, the table of *Household Tenure* is presented below (**Table 5.5**). On the left of this table is the ROP data which represents survey responses from the EASEL population. On the right of this table are results from the CHAIRS model. Data inputted into the CHAIRS model as at 2001 was originally created via microsimulation using the 2001 Census as discussed in the previous chapter (**Chapter 4**). As a result, it is more representative of the actual population size. If the ROP is to be used as a validation dataset, the counts in both tables must be comparable. For this reason, reweighting techniques are used to increase the ROP sample to match the total population size represented in the CHAIRS model. This is discussed in detail in the section to follow.

	Owners		**Private Renters**		**Council Renters**	
LLSOA	**ROP**	**CHAIRS**	**ROP**	**CHAIRS**	**ROP**	**CHAIRS**
E01011339	0	458	1	184	7	786
E01011341	3	620	9	141	22	412
E01011343	22	673	2	104	5	402
E01011345	10	449	6	162	8	513
E01011346	0	422	2	138	7	797
E01011616	19	537	14	135	16	618
E01011620	10	434	3	129	13	473
E01011624	4	330	7	177	18	815
E01011656	15	523	9	117	20	485
E01011657	20	425	10	111	12	625
E01011658	7	296	10	141	34	821
E01011659	14	486	10	104	12	599
E01011660	17	422	5	126	23	622
E01011661	11	450	5	134	21	822
E01011662	3	241	11	95	11	748
E01011663	2	229	6	93	27	664
E01011664	15	404	8	137	13	606
E01011665	17	648	3	177	6	361
E01011666	27	977	0	157	1	66
E01011667	5	225	5	77	23	669
Total	**221**	**9249**	**126**	**2639**	**299**	**11904**

Table 5.5 Counts of Housing Tenure using original ROP and CHAIRS results

Also, notice in the ROP data that there are some areas where the population is not represented. For example, in *E01011339* there are no households recorded as owners although according to the 2001 Census there are owners recorded. Similarly, the 2005/2006 CHAIRS results show 458 owners in this area. Even when the reweighting is applied to the ROP table to systematically increase the counts based on the observed trends, owners will not be represented in this case due to the value zero recorded. This an obvious disadvantage which results in some lost of detail. It can be observed across other household attributes such as age to a limited degree. The ethnicity variable, however, is peculiar where zero values are concerned.

Ethnic segregation is one of the unique features of the EASEL district. Representing this is therefore important in the context of the CHAIRS model. **Table 5.6** illustrates the results from the ROP and the CHAIRS model. According to both data sets, there are few non-whites in the EASEL district, however, the ROP shows no non-whites in many areas. Therefore, due to the absence of ethnicity data in some LLSOAs, discretion is needed when drawing conclusions based on the reweighted figures which will show many areas with 100% Whites not least because of the sample biases mentioned earlier. Notice that there is no sub-categorisation of the non-white group due to the low counts across the individual ethnicity classes.

	White		Non-White	
LLSOA	**ROP**	**CHAIRS**	**ROP**	**CHAIRS**
E01011339	8	1251	0	177
E01011341	34	1089	0	84
E01011343	29	1087	0	92
E01011345	24	993	0	131
E01011346	9	1248	0	109
E01011616	49	1153	0	137
E01011620	25	967	1	69
E01011624	28	1166	1	156
E01011656	43	1035	1	90
E01011657	42	1020	0	141
E01011658	47	1134	4	124
E01011659	36	1084	0	105
E01011660	44	1081	1	89
E01011661	37	1228	0	178
E01011662	24	991	1	93
E01011663	34	871	1	115
E01011664	36	1058	0	89
E01011665	26	1096	0	90
E01011666	28	1094	0	106

E01011667	33	893	0	78
Total	**636**	**21539**	**10**	**2253**

Table 5.6 Tables of Ethnicity counts using original ROP data and CHAIRS results

Similar tables are created for accommodation type and age. Tenure, ethnicity, accommodation type and age are emphasised because they are the four validation variables used. These variables give an idea of the demographic structure of the study area as well as the distribution of households. The variables have also been chosen as they are all represented in the ROP micro data and the CHAIRS model results. They are the variables used to compare counts from the ROP micro dataset with the results of the CHAIRS model.

5.2.5 Preparing the Validation Dataset for Use: Reweighting the ROP

The ROP dataset is reweighted using the 2001 Census. The 2001 Census is used since there are no other datasets available for the 2005/2006 period which would give more accurate trends and counts of households in the EASEL area. Surveys such as the BHPS may offer more recent data but such a survey represents a sample of the population in areas other than EASEL and therefore would not capture the full population distribution in the EASEL district. The disadvantage of using the 2001 Census, however, is the assumption that the population distribution in the EASEL district has not changed since 2001. In lieu of a better alternative, trends from the 2001 Census are used.

The process of reweighting is made up of three functions. First reweighting factors must be calculated then tables of weights are constructed for each validation variable in order to preserve population distribution trends, finally these weights are used to augment the ROP micro data.

5.2.6 Generating the Reweighting Factor

Reweighting is a precursor to validation, it ensures that the ROP micro dataset can be easily compared to the results of the CHAIRS simulation. Using the validation variables; tenure, ethnicity, accommodation type and age, the ROP dataset is aggregated for each of the variables by LLSOA.

Notice the 20 LLSOAs represented in the table below (**Table 5.7**). These are the only LLSOAs from which responses were obtained in the ROP dataset, for this reason, though the reweighting factors generated from this data will be applied to the entire EASEL district, validation can only be performed on these 20 LLSOAs. In total, there are 55 LLSOAs in the EASEL district, by validating against only 20 LLSOAs it is assumed that if the CHAIRS results fit the ROP data well in these areas, then a similar result is likely throughout EASEL.

LLSOA	Owners	Council Renters	Private Renters	Total
E01011339	0	7	1	8
E01011341	3	22	9	34
E01011343	22	5	2	29
E01011345	10	8	6	24
E01011346	0	7	2	9
E01011616	19	16	14	49
E01011620	10	13	3	26
E01011624	4	18	7	29
E01011656	15	20	9	44
E01011657	20	12	10	42
E01011658	7	34	10	51
E01011659	14	12	10	36
E01011660	17	23	5	45
E01011661	11	21	5	37
E01011662	3	11	11	25
E01011663	2	27	6	35
E01011664	15	13	8	36
E01011665	17	6	3	26
E01011666	27	1	0	28
E01011667	5	23	5	33
Total	**221**	**299**	**126**	**646**

Table 5.7 Aggregated data based on Housing Tenure using the original ROP

At the end of the reweighting process, total numbers of, for example, different ethnic groups, will be the same as the 2001 census figures, however, the geographic distribution by OA will be determined by the distributions in the ROP. The total numbers in each OA will match the model's predictions. Reweighting this dataset is a two step process. First the general trends in the ROP are compared to the trends in the 2001 Census; the proportions of residents living in owner-occupation, council housing and private rentals in the ROP are compared to the proportions of residents aggregated for the same variables in the census. If the ROP dataset is to be augmented to mirror the trends in the census data, then these proportions should be equal. More practically, using **Table 5.7** above, there are 221 households living in owner occupation, this equates to 34.2105% of the EASEL residents ((221÷646)*100). However, observations within the 2001 Census show that 36.9112% of EASEL residents live in owner-occupation therefore a reweighting factor is needed to match the proportions in the ROP data with the proportions in the census. In this example, the reweighting factor is 1.0789; 36.9112% ÷ 34.2105%. **Tables 5.8 to 5.11 (Continued)** show the resultant reweighting factors for each category of the four validation variables.

Tenure (%)	Owner-occupation	Council Housing	Private Renting
Census	36.9112	50.41563	12.673179
ROP	34.2105	46.28483	19.504644
Reweight Factor	1.0789	1.0892	0.6498

Table 5.8 Reweighting Factors for Tenure variable

Accommodation Type (%)	Detached	Semi Detached	Terrace	Purpose Built Flat
Census	3.329925	36.56362	35.68263	21.25907863
ROP	15.78947	1.083591	16.56347	59.75232198
Reweighting Factor	0.2109	33.743	2.1543	0.3558

Table 5.9 Reweighting Factors for Accommodation Type variable

Accommodation Type (%)	Flat Converted	Maisonette, Mobile/Temporary Structure
Census	2.131676	1.033063
ROP	3.250774	3.560372
Reweighting Factor	0.6557	0.2902

Table 5.9 (Continued) Reweighting Factors for Accommodation Type variable

Ethnicity (%)	White	Non-White
Census	86.3115	13.6884
ROP	98.4520	1.5479
Reweighting Factor	0.8767	8.8428

Table 5.10 Reweighting Ethnicity variable

Age (%)	<=19	20_24	25_29	30_34	35_39	40_44	45_49
Census	1.63339	5.545163	8.8454558	11.556610	10.855786	9.088370	7.5303643
ROP	1.23839	5.263158	7.4303406	6.1919504	10.216718	8.823529	10.681114
Reweighting Factor	1.319	1.0536	1.1905	1.8664	1.0626	1.03	0.705

Table 5.11 Reweighting Age Category variable

Age (%)	50_54	55_59	60_64	65_69	70_74	>=75
Census	8.00503	6.469356	6.6229234	6.393969007	6.550328075	10.9032528
ROP	5.57276	9.133127	7.120743	6.965944272	7.120743034	14.2414861
Reweighting Factor	1.4365	0.7083	0.9301	0.9179	0.9199	0.7656

Table 5.11 (Continued) Reweighting Age Category variable

Based on these tables, there are obvious similarities and differences between the ROP data and the 2001 census. For example, **Table 5.8** shows the results of the tenure variable. Here there are slightly less owners reported by the ROP and slightly less council tenants in the same dataset while there is are more private renters in this dataset. Though the cause of these differences is not clear, this may be due to the change in the population between 2001 and 2005. This is true for each of the four variables used and therefore differences will be realised in subsequent tables. In **Table 5.8**, the assumption may be made that the over-representation in private renters could be linked to the over-representation of White British respondents noted in **Table 5.10**. While the generally low level of minority respondents may be linked to the lack of interest in engagement activities by minority groups. This latter assumption is particularly important because of the lack of Asians represented in the ROP dataset. As mentioned previously, to remedy this, ethnicity is divided into two categories; White and Non-White, in this way the Non-White group can be reweighted using the census proportions. This group includes Asians, who are not represented in the ROP dataset.

In essence this is the first part of the reweighting process. Reweighting factors have been created in **Tables 5.8** to **5.11 (Continued)** to ensure that the trends in the 2001 Census are reflected in the ROP dataset.

5.2.7 Using the Reweighting Factor

With the reweighting factors generated for each category of each variable in the ROP, these values are used to generate tables of weights for the entire ROP dataset by LLSOA. The tables of weights from the ROP dataset are then used to augment the original ROP data to match the actual population counts in the EASEL district. The weights are first used to ensure that population levels are retained as discussed earlier and to ensure that population counts are the same for each OA in both the augmented ROP and the actual population of EASEL generated by the simulation model.

The tenure table is used in **Tables 5.12** and **5.13** to illustrate how the table of ROP weights is calculated. Using the reweighting factor as previous calculated and repeated in **Table 5.12**, each value in the original ROP data is multiplied by the corresponding reweighting factor for the variable category. For example, in **Table 5.13**, the council housing value of *E01011339* is multiplied by the reweighting factor 1.0892 resulting in a new weight for the council housing category of 7.6244; 7 * 1.0892. In a similar way, the private renting category of *E01011343* can be extended to create a new weight for this category of 1.2996; 2 * 0.6498.

Tenure (%)	Owner-occupation	Council Housing	Private Renting
Census	36.9112	50.41563	12.673179
ROP	34.2105	46.28483	19.504644
Reweight Factor	1.0789	1.0892	0.6498

Table 5.12 Reweighting Factors for Tenure variable

	Ownership		Council Housing		Private Renting	
LLSOA	**Original ROP**	**Extended Weights**	**Original ROP**	**Extended Weights**	**Original ROP**	**Extended Weights**
E01011339	0	0	7	7.6244	1	0.6498
E01011341	3	3.2367	22	23.9624	9	5.8482
E01011343	22	23.7358	5	5.446	2	1.2996
E01011345	10	10.789	8	8.7136	6	3.8988
E01011346	0	0	7	7.6244	2	1.2996
E01011616	19	20.4991	16	17.4272	14	9.0972
E01011620	10	10.789	13	14.1596	3	1.9494
E01011624	4	4.3156	18	19.6056	7	4.5486
E01011656	15	16.1835	20	21.784	9	5.8482
E01011657	20	21.578	12	13.0704	10	6.498
E01011658	7	7.5523	34	37.0328	10	6.498
E01011659	14	15.1046	12	13.0704	10	6.498
E01011660	17	18.3413	23	25.0516	5	3.249
E01011661	11	11.8679	21	22.8732	5	3.249
E01011662	3	3.2367	11	11.9812	11	7.1478
E01011663	2	2.1578	27	29.4084	6	3.8988
E01011664	15	16.1835	13	14.1596	8	5.1984
E01011665	17	18.3413	6	6.5352	3	1.9494
E01011666	27	29.1303	1	1.0892	0	0
E01011667	5	5.3945	23	25.0516	5	3.249
Total	**221**	**238.4369**	**299**	**325.6708**	**126**	**81.8748**

Table 5.13 Original versus Extended Tenure Weights

This process is repeated for the accommodation type, age and ethnicity. Finally, the table of detailed ROP weights is used to generate population counts using the total count of households generated by the model for each LLSOA. **Table 5.14** illustrates this process. For each LLSOA, each weight in the Extended ROP Weights table is divided by the total weight then multiplied by the EASEL count to generate an actual count of households. That is, for LLSOA *E01011339*, the actual household count for council housing is calculated using the formula (7.6244 ÷ 8.2742) * 1421. In this way, the EASEL counts for each LLSOA reported by the model can now be matched with the Augmented ROP Counts. The Extended ROP Weights tables for each of the four validation variables are used to augment the ROP counts in this way. Note that these Augmented ROP Counts tables need to be generated for each model run as the number of households in each LLSOA will vary from one model run to another.

Extended ROP Weights (Tenure)					**Augmented ROP Counts (Tenure)**			
LLSOA	**Ownership**	**Council Housing**	**Private Renting**	**Total Weight**	**Ownership**	**Council Housing**	**Private Renting**	**Total**
E01011339	0	7.6244	0.6498	8.2742	0	1309.404	111.5958	1421
E01011341	3.2367	23.9624	5.8482	33.0473	114.8853	850.5353	207.5794	1173
E01011343	23.7358	5.446	1.2996	30.4814	898.6173	206.1809	49.20176	1154
E01011345	10.789	8.7136	3.8988	23.4014	522.8203	422.2492	188.9305	1134
E01011346	0	7.6244	1.2996	8.924	0	1148.274	195.7264	1344
E01011616	20.4991	17.4272	9.0972	47.0235	567.149	482.1586	251.6924	1301
E01011620	10.789	14.1596	1.9494	26.898	414.7456	544.3165	74.9379	1034
E01011624	4.3156	19.6056	4.5486	28.4698	195.2417	886.9754	205.7829	1288
E01011656	16.1835	21.784	5.8482	43.8157	422.9102	569.2635	152.8262	1145
E01011657	21.578	13.0704	6.498	41.1464	610.425	369.7516	183.8234	1164
E01011658	7.5523	37.0328	6.498	51.0831	191.1615	937.363	164.4754	1293
E01011659	15.1046	13.0704	6.498	34.673	514.9147	445.569	221.5163	1182
E01011660	18.3413	25.0516	3.249	46.6419	461.2665	630.0242	81.7093	1173
E01011661	11.8679	22.8732	3.249	37.9901	435.478	839.3039	119.2181	1394
E01011662	3.2367	11.9812	7.1478	22.3657	158.4653	586.5863	349.9484	1095
E01011663	2.1578	29.4084	3.8988	35.465	59.32201	808.4926	107.1854	975
E01011664	16.1835	14.1596	5.1984	35.5415	520.4547	455.3669	167.1784	1143
E01011665	18.3413	6.5352	1.9494	26.8259	811.571	289.1714	86.2576	1187
E01011666	29.1303	1.0892	0	30.2195	1161.568	43.43176	0	1205
E01011667	5.3945	25.0516	3.249	33.6951	156.2551	725.6355	94.10935	976
Total				**645.982**				**23781**

Table 5.14 Extended ROP Weights versus Augmented ROP Counts

In effect, the reweighting process ensures that the ROP micro data can be matched with the model results in such a way that the counts are comparable. This systematic exercise is necessary so as to ensure that the ROP is not arbitrarily augmented but that the geographical trends in the data are preserved.

5.2.8 Acknowledging the Potential for Errors

An absolutely accurate result will almost never be generated in the reweighting process. Both the ROP and the 2001 Census are estimates. Furthermore, the census is manipulated to ensure that data protection laws are adhered to and privacy clauses are maintained. Note that the census data used is not the original raw data and it is likely that within this dataset some LLSOAs originally contained zero values as in the micro data of the ROP. Due to the extensive cleaning process which is applied to the census, problems such as these are eliminated. In using the raw micro data from the ROP, however, these problems are still present. Also, it is important to recognise that the levels within the 2001 Census are represented in the Augmented ROP micro data. If the reweighting factors are correct, then the levels realised in the tenure, accommodation, ethnicity and age variables of the census will be represented in the ROP dataset. The only way to improve upon this is to have a more accurate validation data set as was previously discussed.

5.2.9 Measuring Error

In order to usefully analyse the differences between the ROP data and the CHAIRS results, some goodness of fit measure must be used. In this way, the number/proportion of errors can be identified. For this simulation, the TAE is used. The TAE is a simple test statistic used to measure the difference between the observed and estimated population counts (Williamson *et al.*, 1998). It is a goodness of fit indicator and is calculated by summing the absolute differences between corresponding cell counts as illustrated in the formulae:

$$TAE = \sum_{ij} \left| U_{ij} - T_{ij} \right|$$

Where,

U_{ij} is the observed count for row *i* in column *j*, and

T_{ij} is the expected count for row *i* in column *j*.

The TAE is an alternative to parametric tests which are noted to be inappropriate due to the high level of autocorrelation between demographic attributes. Other nonparametric tests are also of limited use as they do not follow a generic process that can be universally applied (WILLIAMSON, P. *et al*., 1998). Though other competing techniques may be used in this calibration validation process, for example, confusion matrices, the TAE and SAE error statistics are thought to be sufficiently robust is ascertaining the differences in population counts between the validation dataset and the results of the CHAIRS simulation.

A TAE of zero is the ideal result. This indicates that there is no error between the observed and estimated population counts. However, with the alterations in datasets for data protection and privacy reasons and with assumptions governing some methodology in the model design, a TAE value of zero is unlikely. In general, the TAE has a maximum value of twice the total population count. In the case of the CHAIRS simulation, a household incorrectly placed affects the error in both the location it is placed in, and the location in which it should be placed.

The object of this validation exercise is to find the model which generates the least errors. Unlike the TAE, which is an absolute count, the SAE (SAE) allows for comparisons to be made across different tables or variables (KONGMUANG, C. *et al*., 2005). It is calculated by dividing the TAE by the population count for each table. As there are four validation variables used in this model, it is the average SAE across these four variables that is generated. In this way the performance of each rule-set can be ranked with the rule-set generating the least amount of error viewed as the one with the best fit to the actual real-world data represented in the augmented ROP dataset. **Table 5.15** is a condensed list of SAE statistics. It is an example and therefore does not represent actual results. Based on the results shown, rule-set 5 produces the lowest average error therefore meaning that rule-set 5 produces the best fit to the real-world; it is the best performing model out of the seven models presented. It may be argued that rule-set 2 produces a lower ethnicity and accommodation type error than rule-set 5 suggesting that other rule-sets may have the advantage of producing lower errors when individual variables are considered. However, though different combinations of lowest error can be chosen, in this case, one type of error is not favoured over the other; tenure is not more important than age rather it is the overall resultant average error that is most important.

Rule-set	Tenure	Accommodation Type	Ethnicity	Age	Average Error
1	0.365725	1.043508	0.233255	0.63556	0.569512
2	0.375245	1.049754	0.217666	0.633401	0.569016
3	0.390697	1.056037	0.242531	0.612235	0.575375
4	0.386299	1.040579	0.229281	0.633427	0.572397
5	0.332011	1.057996	0.230843	0.638706	0.564889
...	...	...	...	...	...
126	0.369273	1.121781	0.219408	0.632241	0.585676
127	0.371696	1.131229	0.221564	0.645184	0.592418

Table 5.15 SAE for Rule-set 1-5

5.2.10 Using the Validation Dataset in an Example of the Calibration/Validation Process

With the ROP micro data augmented to represent actual counts in the EASEL area and a goodness of fit measure defined, the calibration and validation processes can be described more practically in an example. The model is executed using each of the 127 rule-set combinations originally identified in **Appendix C**. Using the 2005/2006 CHAIRS model results, the individual-level results are aggregated so that detailed tables of counts for each LLSOA can be created for each of the validation variables; tenure, accommodation type, ethnicity and age. An example of the tenure output from the model is shown below (**Table 5.16**).

LLSOA	Ownership	Council Housing	Private Renting	Total
E01011339	489	862	163	1514
E01011341	541	496	110	1147
E01011343	562	520	86	1168
E01011345	464	565	119	1148
E01011346	449	813	204	1466
...	...	...	...	...
E01011662	303	735	140	1178
E01011663	280	662	112	1054
E01011664	402	674	134	1210
E01011665	594	449	117	1160

E01011666	704	315	91	1110
E01011667	264	651	126	1041
Total	**9048**	**12847**	**2820**	**24715**

Table 5.16 Output of CHAIRS Model using rule-set 1

This table can then be compared to the corresponding Augmented ROP counts table for the purpose of validation. The corresponding table is shown in part in **Table 5.17**.

LLSOA	Ownership	Council Housing	Private Renting	Total
E01011339	0	1395.101	118.8994	1514
E01011341	112.3388	831.6829	202.9783	1147
E01011343	909.5191	208.6823	49.79866	1168
E01011345	529.2748	427.4622	191.263	1148
E01011346	0	1252.507	213.4932	1466
...	...	...	...	...
E01011662	170.4768	631.049	376.4742	1178
E01011663	64.12861	874.0012	115.8702	1054
E01011664	550.9625	482.0595	176.978	1210
E01011665	793.1107	282.5938	84.29555	1160
E01011666	1069.992	40.00768	0	1110
E01011667	166.6615	773.9617	100.3769	1041
Total				**24715**

Table 5.17 Augmented ROP Counts for rule-set 1

With these two tables available, that is, the model outputs (observed population counts) and the Augmented ROP micro data (estimated population counts), an error statistic can be calculated for the purpose of validation. Using the TAE goodness of fit measure, the sum of the absolute differences between corresponding variable values in each LLSOA can be calculated. While the absolute differences represent the number of errors in each category for a given variable, summing the absolute differences in the **Table 5.18** reveals that when the tenure variable is analysed in rule-set 1, there are 6457 incorrectly placed households out of a possible 24715 total households. Each misplaced household creates a discrepancy in the correct location and the location in which it was placed, therefore there are 12914 errors (6457*2).

LLSOA	Ownership	Council Housing	Private Rental
E01011339	489	533	44
E01011341	429	336	93
E01011343	348	311	36
E01011345	65	138	72
E01011346	449	440	9
...	...	...	...
E01011662	133	104	236
E01011663	216	212	4
E01011664	149	192	43
E01011665	199	166	33
E01011666	366	275	91
E01011667	97	123	26

Table 5.18 Absolute Differences between **Table 5.16** and **5.17**

These two values are important as they are included in a table of errors for all rule-sets and are further used in the calculation of the SAE. The TAE statistic is calculated for each validation variable; accommodation type, ethnicity and age and in turn for each of the 127 rule-sets. **Table 5.19** is a condensed list of TAE statistics.

Rule-set	Tenure	Accommodation Type	Ethnicity	Age	Total Households
1	9038.903	25790.311	5764.903	15707.86	24715
2	9258.431	25900.582	5370.465	15627.89	24673
3	9867.041	26670.225	6125.12	15462.01	25255
4	9582.156	25811.571	5687.309	15712.16	24805
5	8161.153	26006.61	5674.351	15700.02	24581
...	...	...	...	...	...
126	8825.246	26809.452	5243.632	15109.93	23899
127	8907.313	27108.769	5309.57	15461.19	23964

Table 5.19 TAE for condensed rule-set list

Likewise, the SAE allows for comparison across all rule-sets is detailed in **Table 5.20** below.

Rule-set	Tenure	Accommodation Type	Ethnicity	Age	Average SAE
1	0.320578	1.067922	0.251202	0.616413	0.564029
2	0.326088	1.073338	0.24509	0.614168	0.564671
3	0.315919	1.076713	0.254904	0.612342	0.564969
4	0.322209	1.069649	0.250162	0.617976	0.564999
5	0.348378	1.059714	0.24818	0.609381	0.566413
...	...	...	...	...	...
126	0.369273	1.121781	0.242348	0.632241	0.591411
127	0.371696	1.131229	0.243841	0.645184	0.597987

Table 5.20 SAE for condensed rule-set list

Based on these example results, rule-set 1 is chosen as the best performing rule-set as it produces the lowest SAE statistic. A similar process is executed when the additional parameters are explored as mentioned in **Section 5.3.3**. This is the methodology that will be used to calibrate and validate the CHAIRS model.

5.3 Calibrating the CHAIRS Model

Having discussed the methodology to be used for calibration and validation, the results of the CHAIRS simulation model will be explored. In this way, parameters can be altered with the aim of finding the combination of parameters which best fit the validation dataset. Initially, each of the 127 rule-sets will be examined to determine the rule-set which produces the least number of errors on average. Once the best performing rule-set is identified, parameters within the model will be altered systematically in an attempt to further reduce the number of errors generated.

5.3.1 Exploring the Model to Find the Best Rule-set Combination

Having executed the 127 rule-sets, **Table 5.21** shows the top 20 performing rule-sets while **Appendices D** and **E** present a complete list of the 127 rule-sets ranked in order of best performance.

Rank	Rule-set	Rule Descriptions	SAE
1	61	Ethnicity, Socio Economic Status, Transport Routes	0.564029
2	81	Ethnicity, Socio Economic Status, Transport Routes, Schools	0.564671
3	58	Schools, Socio Economic Status, Transport Routes	0.564969
4	42	Rooms, Socio Economic Status, Transport Routes	0.564999
5	52	Output Area, Socio Economic Status, Transport Routes	0.566413
6	91	Rooms, Transport Routes, Schools, Socio Economic Status	0.566648
7	11	Rooms, Socio Economic Status	0.567477
8	20	Socio Economic Status, Schools	0.568008
9	54	Ethnicity, Socio Economic Status, Schools	0.568682
10	27	Socio Economic Status, Transport	0.568827
11	82	Ethnicity, Socio Economic Status, Transport Routes, Output Areas	0.568861
12	115	Ethnicity, Socio Economic Status, Transport Routes, Schools, Output Area	0.569063
13	18	Output Area, Transport Routes	0.56913
14	103	Socio Economic Status, Transport Routes, Schools, Output Areas, Rooms	0.569178
15	110	Ethnicity, Socio Economic Status, Transport Routes, Schools, Rooms	0.569255
16	33	Transport, Output Areas, Rooms	0.569502
17	69	Ethnicity, Transport Routes, Output Areas, Rooms	0.569736
18	38	Ethnicity, Socio Economic Status, Rooms	0.569744
19	72	Ethnicity, Socio Economic Status, Schools, Rooms	0.570037
20	35	Socio Economic Status, Schools, Rooms	0.570263

Table 5.21 Top 20 performing rule-sets ranked in order of lowest average SAE

Here the average SAE of the four validation variables is used to rank the rule-sets in order of best performance; that is, the model which generates the least number of errors. As shown from the table, rule-set 61 is ranked first with an average SAE of 0.564029. This rule-set is comprised of three rules; *Ethnicity*, *Socioeconomic Status* and *Transport Routes*. In effect, these results indicate that out of the 127 rule-set combinations available, these three rules are able to recreate the population distribution of the EASEL area with the least number of errors. In a similar way, rule-set 81 is ranked second. This rule-set adds the *Schools* rule to rule-set 61. The table continues by detailing the ranking of subsequent rule-set combinations.

Recall that the SAE is used so as to allow comparison from one rule-set to another; however, a look at the actual number of errors generated using the TAE is useful. Using the TAE allows for a clearer view of the actual number of errors generated at

LLSOA level, in this way, more details can be presented. The SAE results used only shows aggregate variation at the EASEL area level. Using the TAE statistics, the graphs below (**Figures 5.2-5.5**) illustrate the performance of a range of rule-sets. In this case model 61 is ranked 1st while models 40 and 94 are ranked 63rd and 127th respectively. The models have been chosen to show the best and worst performances. In these graphical illustrations, the error totals can be observed by LLSOA.

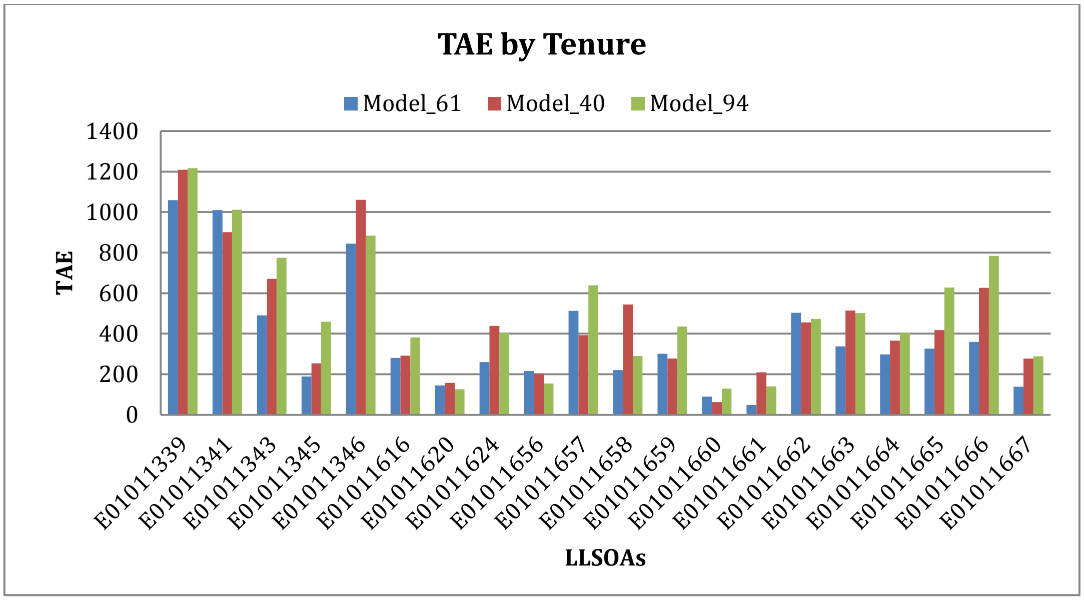


Figure 5.2 Comparing TAE by Tenure for rule-set 61, 40 and 94

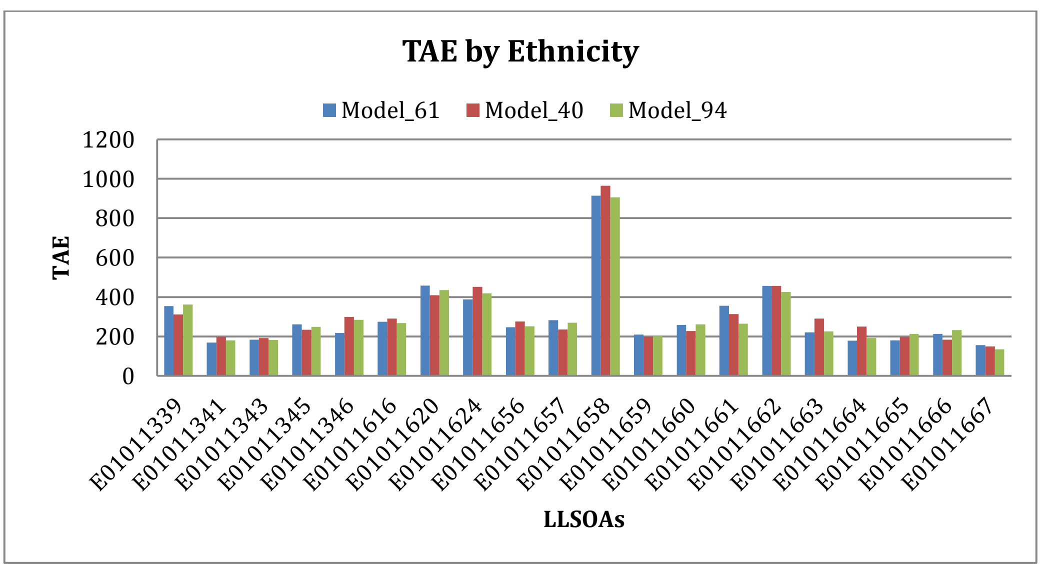


Figure 5.3 Comparing TAE by Ethnicity for rule-set 61, 40 and 94

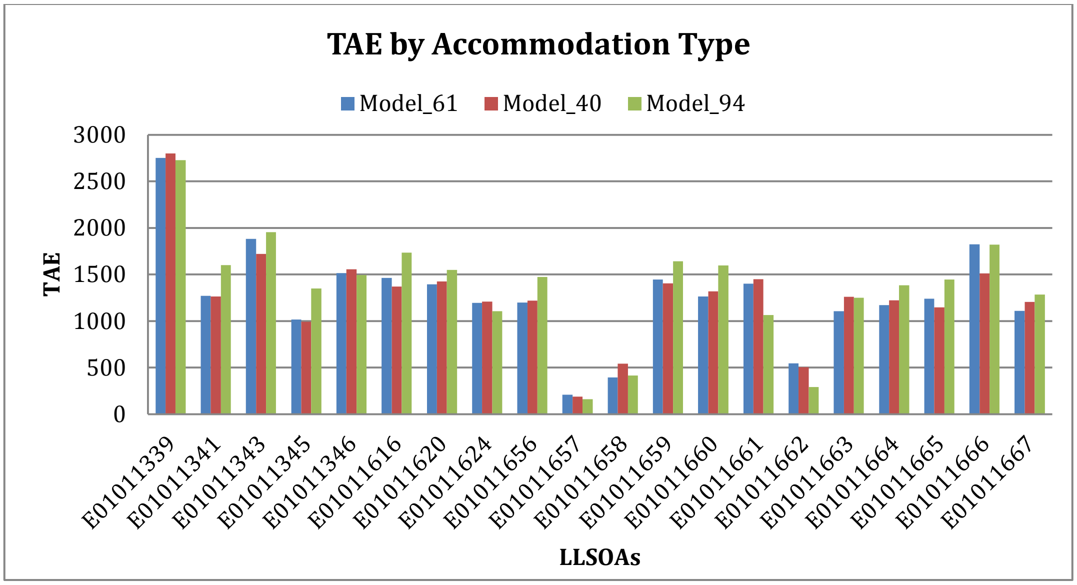


Figure 5.4 Comparing TAE by Accommodation Type for rule-set 61, 40 and 94

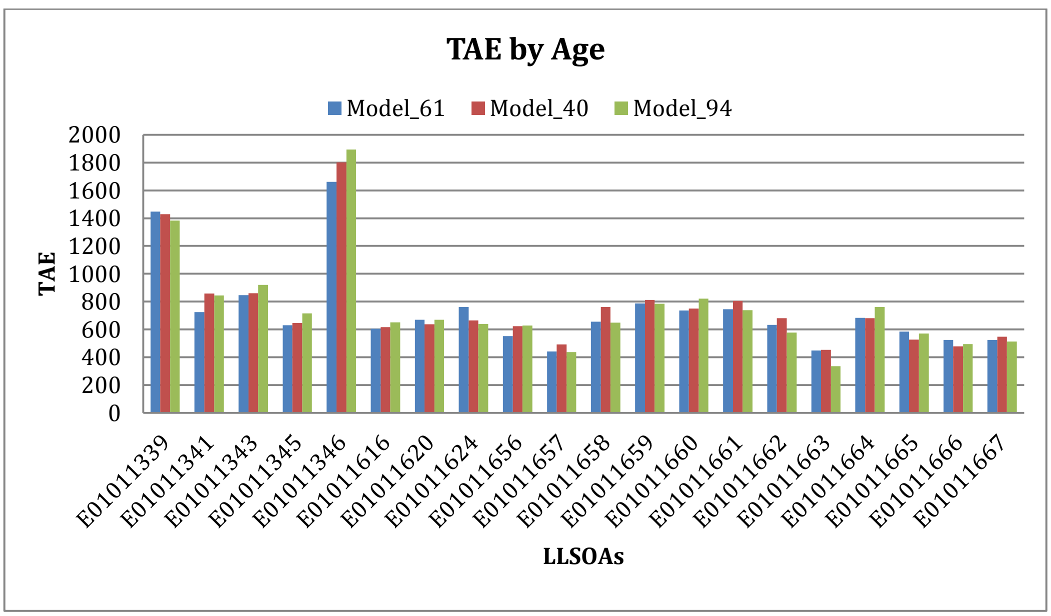


Figure 5.5 Comparing TAE by Age for rule-set 61, 40 and 94

Based on the graphs presented, model 61 produces the lowest number of errors in most LLSOAs when the tenure, accommodation type and age validation variables are observed. Though this may not be the case for the ethnicity variable, it is the total number of errors across these four validation variables that is used to find the model which produces the lowest total number of errors. In this case, model 61 produces the lowest total number of errors and is chosen as the rule-set that produces the best fit to the validation dataset. Note that one anomaly can be observed when the ethnicity data is considered; *E01011658* is significantly underestimated in the ROP reweights as compared to the Census data. This may be linked to the bias in the ROP dataset earlier discussed. Overall, model 61 produces the least number of errors despite the diverse distribution of errors across each rule-set detailed in **Table 5.22** below.

Validation Variable	Rule-set_61	Rule-set_40	Rule-set_94
Tenure	7627.189	9329.125	10123.61
Ethnicity	5976.604	6125.547	5953.663
Accommodation Type	25407	25315.93	27358.83
Age	14665.7	15126.01	15026.39
Total Errors	**53676.49**	**55896.61**	**58462.49**

Table 5.22 Summary of TAE results for rule-sets 61, 40 and 94

Extending these results further, **Table 5.23** compares the results of the top 5 performing rule-set as well as the random allocation of households using the standard deviation as the evaluation statistic. Here, each model was run up to 5 times, in this way the consistency of the chosen rule-set, number 61, could be justified. The table below shows that rule-set 61 generates lower statistics when the standard deviation is concerned while the random allocation of households generates higher standard deviations. This indicates that when rule-set 61 is executed several times, the results are consistently lower than any other rule-set combination while all rule-sets generated lower rates for the standard deviation when compared to the random allocation of households.

Rule-set combination	Tenure	Ethnicity	Accommodation Type	Age
Ethnicity, Socio Economic Status, Transport Routes (61)	92.43	127.15	166.29	74.84
Ethnicity, Socio Economic Status, Transport Routes, Schools (81)	183.90	133.01	154.71	55.31
Schools, Socio Economic Status, Transport Routes (58)	150.60	169.43	178.30	201.04
Rooms, Socio Economic Status, Transport Routes (42)	186.89	141.90	67.61	119.27
Output Area, Socio Economic Status, Transport Routes (52)	165.34	62.408	135.90	292.34
Random allocation to households	678.27	202.18	263.84	332.64

Table 5.23 Multiple model runs for top 5 rule-set combinations and random household allocation using the Standard Deviation based on 5 model runs each

5.3.2 Key Observations and Analysis of Results

Based on these results, some observations can be made. When the TAE of rule-set 61 is compared to the average TAE for all rule-sets, rule-set 61 continues to produce better results. **Figures 5.6-5.9** below illustrate the results.

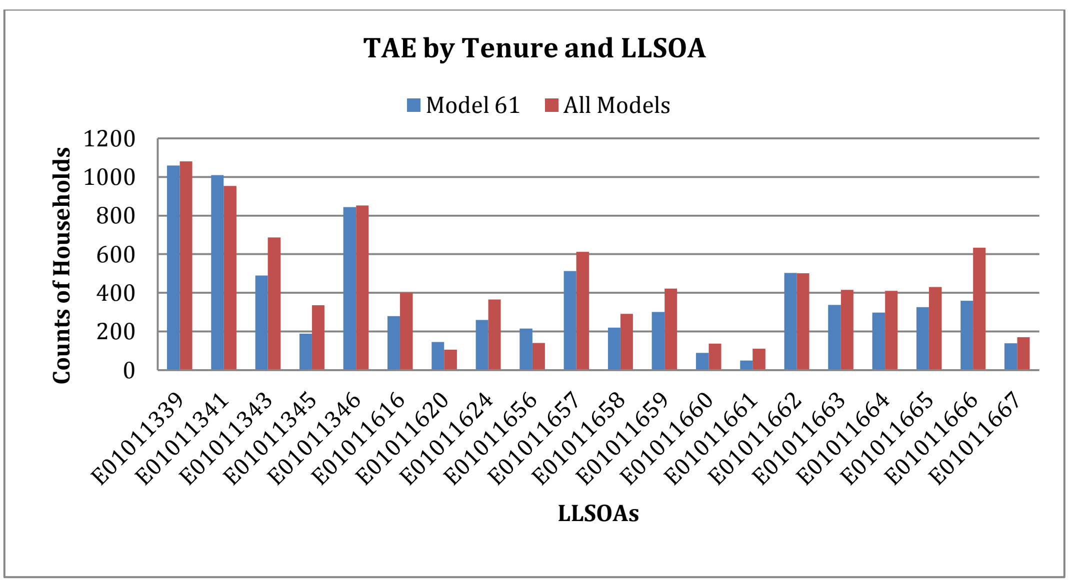


Figure 5.6 TAE by Tenure and LLSOA

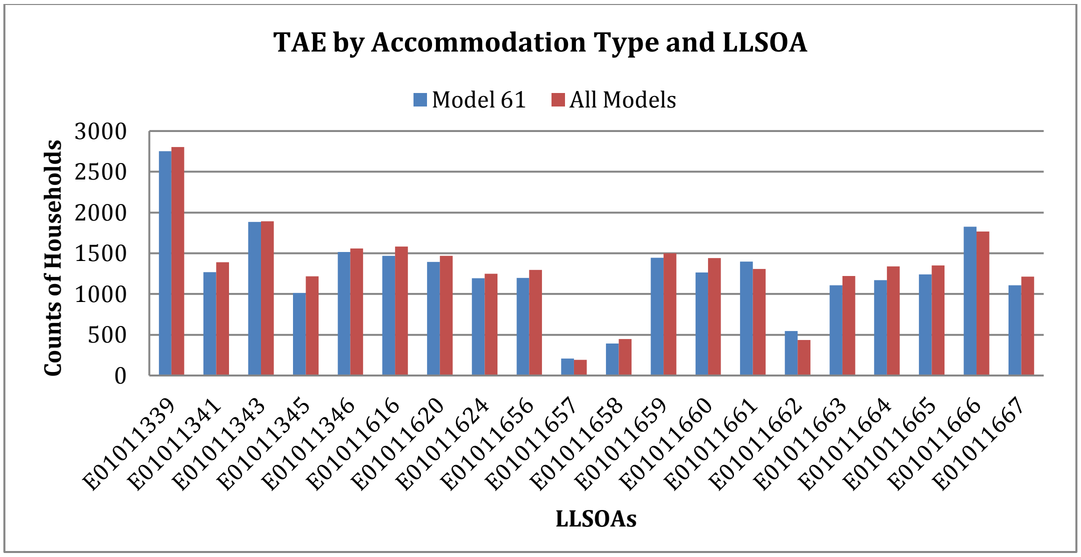


Figure 5.7 TAE by Accommodation Type and LLSOA

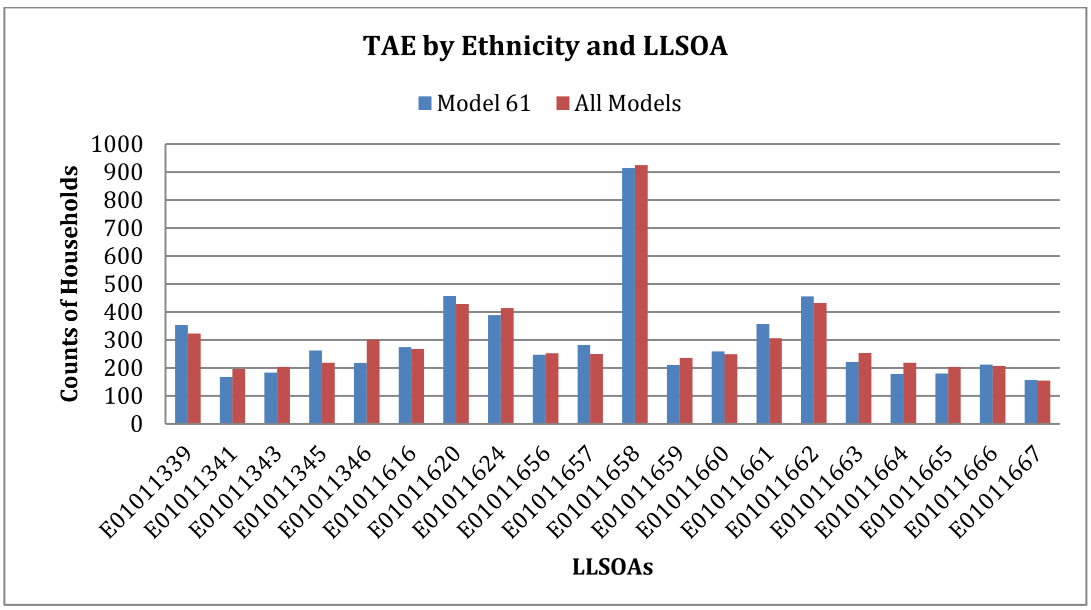


Figure 5.8 TAE by Ethnicity and LLSOA

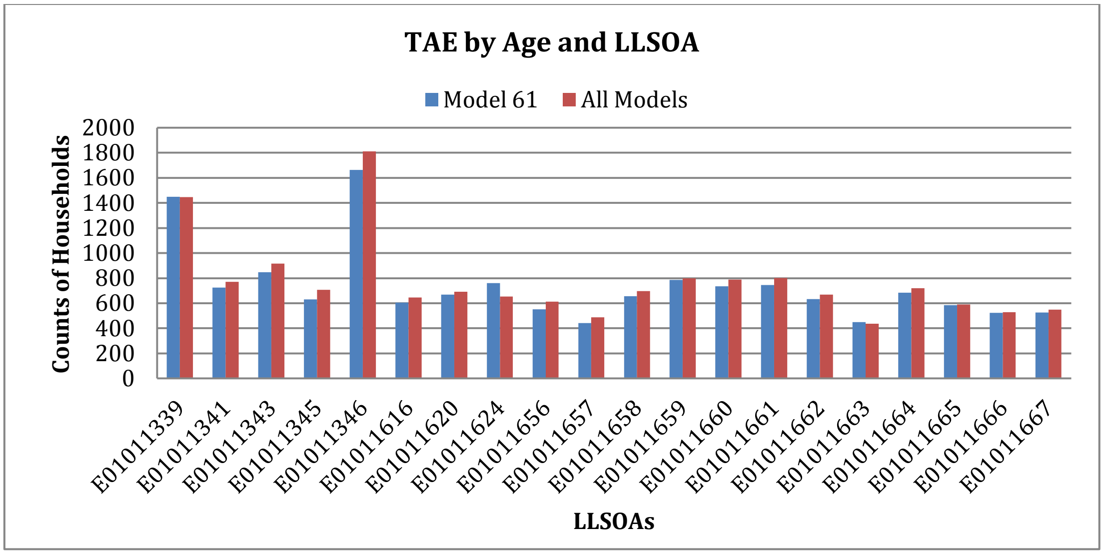


Figure 5.9 TAE by Age and LLSOA

As originally observed, the tenure, accommodation type and age variables perform better than average when the performance of rule-set 61 is compared to the average performance of all rule-sets. Though this is not true for the ethnicity variable, it should be noted that this variable generally produces the least number of errors in the CHAIRS model. Again it must be reiterated that it is the overall lowest number of errors that is used to determine the best performing model.

Another observation that can be made relates to the performance of individual rules. In general when the full complement of results are perused, rule-sets containing parameters based on ethnicity, social status, transport, house size, schools and neighbourhood quality, all contribute to creating EASEL population distributions with the least number of errors. **Figure 5.10a** to **g** below present a graph for each rule parameter showing the significance of individual rules by way of ranking their contribution. Using the full ranking of all rule-sets (**Appendix D, E**), the graphs show the point of time at which each rule contributes. For example, when the rule on familiar neighbourhoods, *Known Areas* (**Figure 5.10e**), only at the rule-set combination ranked in position 60 is this rule first seen. This can be contrasted to **Figure 5.9a** where the ethnicity rule is observed. Here the ethnicity rule is used in the rule-set combination ranked number 1 and all subsequent rule-set indicating that when households try to find houses taking ethnicity into account the error statistic is lower as compared to when the *Known* Areas rule is used. It should be remembered that familiar neighbourhoods in this context defines a list of areas which may be frequented by each household.

The rule on S*ocial Class* ranks consistently high; this is noteworthy as social status is used as a proxy for wealth in the model. In general, households are only able to relocate if they can afford to move and this affordability factor helps to determine the housing tenure chosen (Böheim and Taylor, 2002). It is evident that the transport rule is also important when a new place of residence is to be found. In this case, households attempt to live in areas where public transport is accessible (Rabe and Taylor, 2010). Overall **Figure 5.10a** to **g** below show that *Ethnicity*, *Social Status* and *Transport* are ranked consistently higher than all other rules. The rules based on *OAs*, *Schools* and *Rooms* also fall into higher rankings while the *Known Area* rule is only seen after ranking 60.

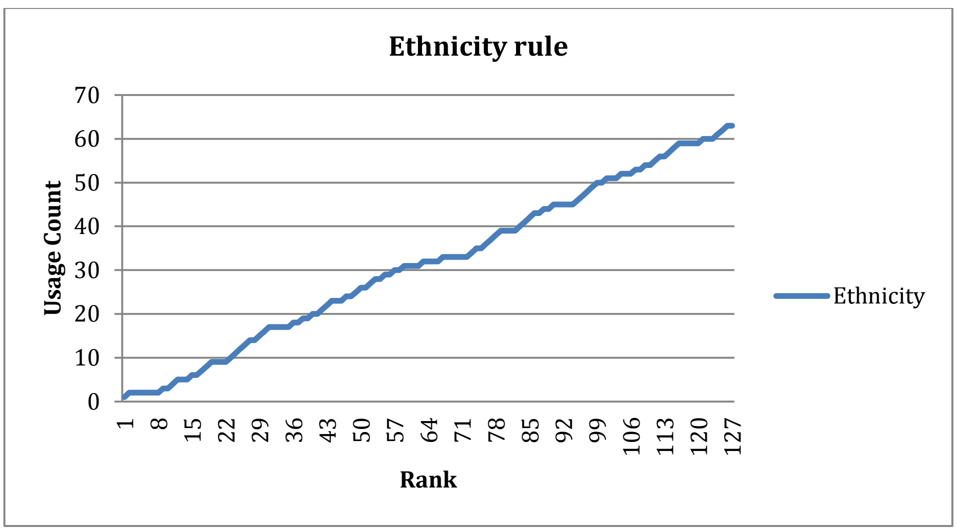


Figure 5.10a Ranking versus usage count for the Ethnicity rule

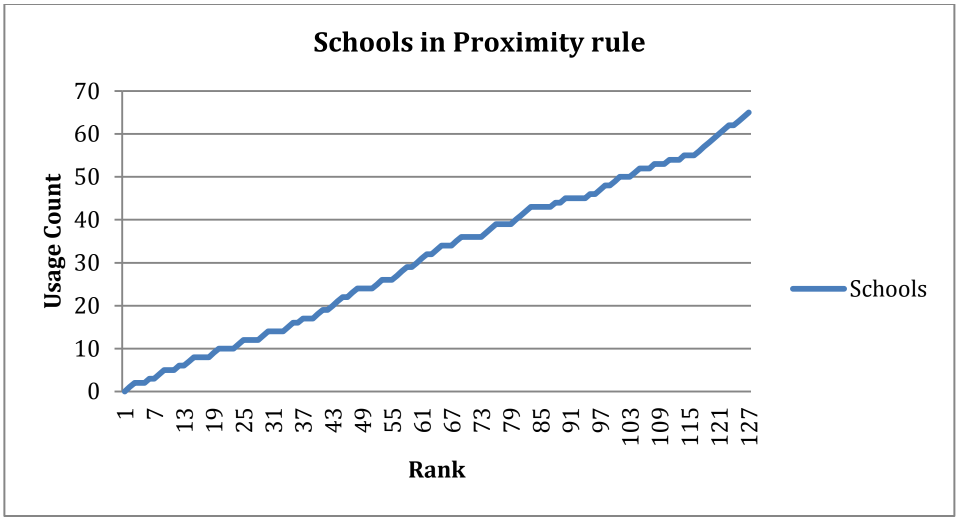


Figure 5.10b Ranking versus usage count for the Schools rule

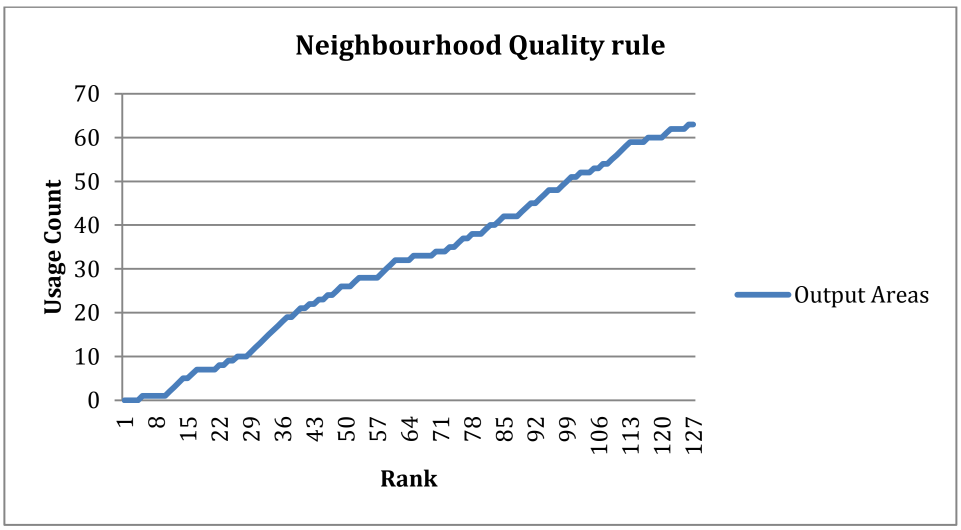


Figure 5.10c Ranking versus usage count for the Neighbourhood Quality rule

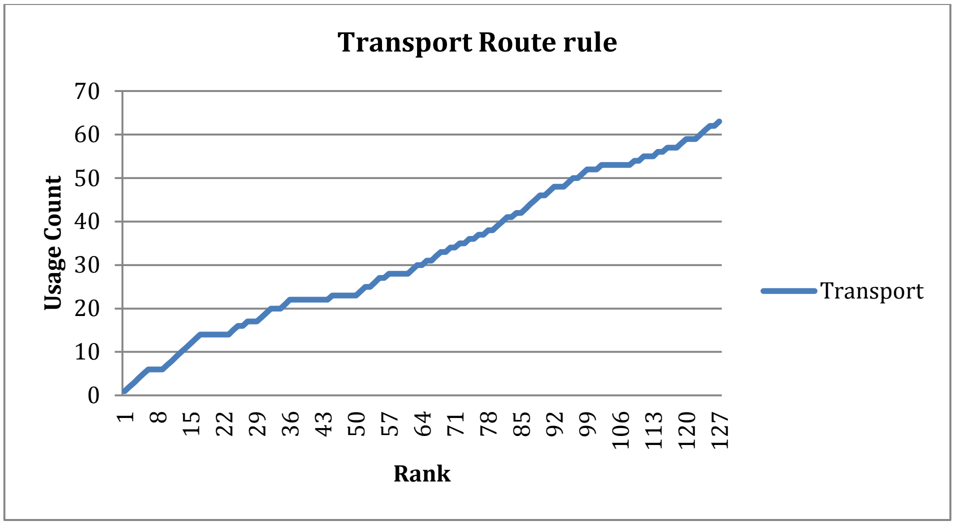


Figure 5.10d Ranking versus usage count for the Transport Route rule

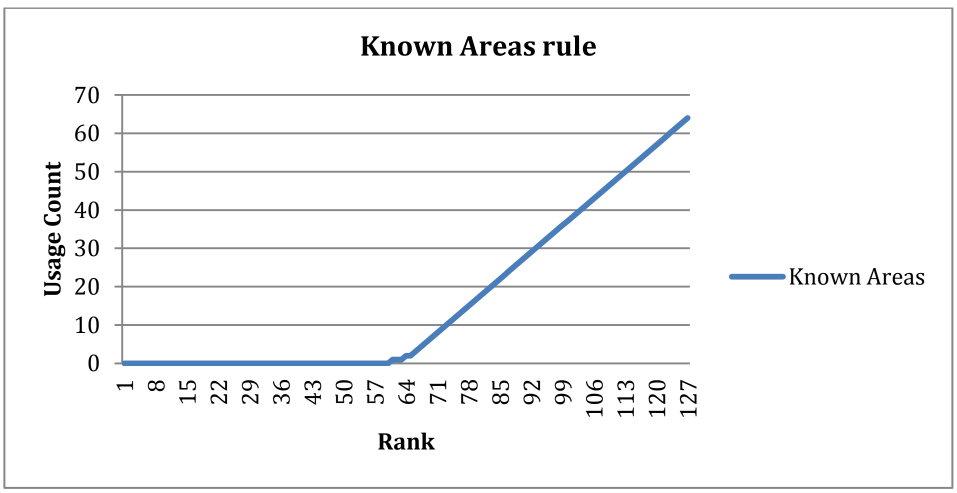


Figure 5.10e Ranking versus usage count for the Known Areas rule

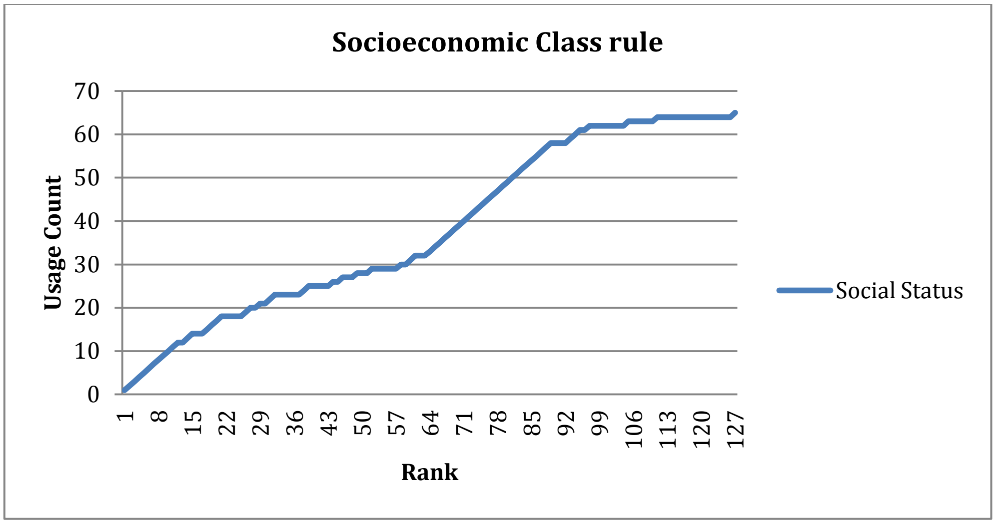


Figure 5.10f Ranking versus usage count for the Socioeconomic Class rule

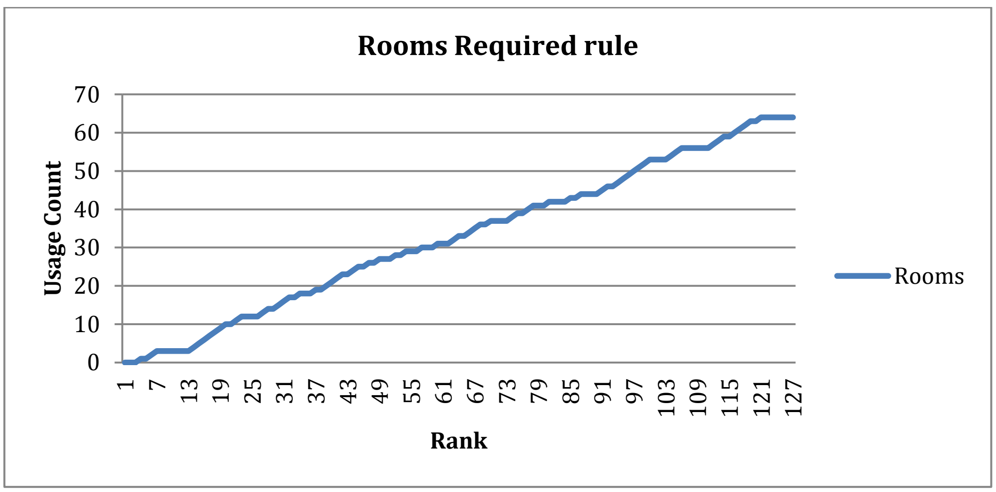


Figure 5.10g Ranking versus usage count for the Rooms Required rule

Analysing the results more generally, **Table 5.24** presents a summary.

Statistical Measure	Ethnicity	Tenure	Age	Accommodation Type
Minimum	4975.47	7605.75	14403.33	24997.10
Maximum	6137.86	10573.45	16459.71	28896.84
Median	**5492.83**	**9133.82**	15274.37	26620.09

Table 5.24 Summary of TAE Results

Across the rule-sets, with a median TAE value 5492, the model appears to best predict the ethnic distribution. Despite the ethnicity biases mentioned earlier, the model still appears to perform better on this variable than others. The ethnicity variable is followed by tenure which has maximum and minimum TAE values of 10573 and 7605 respectively. On average, given there are ~25000 households in the LLSOAs used for validation, these statistics show that there is an error of 11% realised for the ethnicity variable ((5492.83 ÷ (25000*2)) * 100) and an error of 18% realised for the tenure variable ((9133.82 ÷ (25000*2)) * 100). These statistics begin to rise however, when the age and accommodation type variables are analysed. With a maximum TAE of 16459 and a minimum 14403, there is a 31% chance of error recorded for the accommodation type variable. The distribution for age is poorly predicted in this exercise as it reports a 55% average TAE, with maximum and minimum values of 26620 and 24997 respectively.

It is apparent that where ethnicity and housing tenure are concerned, households are being placed in similar locations as recorded in the ROP micro data. The accuracy of this process appears to be reduced when the accommodation type and age variables are considered however. The latter situation is likely due to the fact that none of the behavioural rules govern the way accommodation type decisions are made neither do they include age related decisions. Age is only factored into the decision to move using the '*propensity to move*' statistic (**Section 4.3.3.1**). On the other hand, there are behavioural rules specifically related to ethnicity and housing tenure.

Table 5.25 below is a corresponding summary of TAE percentages, a complete table of results is presented **Appendix D, E**.

Statistical Measure	Ethnicity	Tenure	Age	Accommodation Type
Minimum	11.57	15.8	51.62	29.46
Maximum	13.08	21.13	58	33.12
Median	12.40	18.58	55.55	31.45

Table 5.25 Percentage Error by Validation Variable

It would be useful to compare these statistics with similar statistics generated from other ABMs. However, here is where there is a scarcity of information in the literature related to housing ABMs; unless researchers are willing to openly discuss the overall performance of ABMs, such comparisons will never be possible. However, if the percentage error is viewed at the super output area level then the percentage error is significantly reduced. Using **Figure 5.11** below, for each LLSOA the error statistics appear small with a maximum error of 8%% (0.08*100) reported for the accommodation type variable and a minimum error of less than 1% for the tenure variable. It is evident that the aggregated error results hide much of the variation evident at lower geographies.

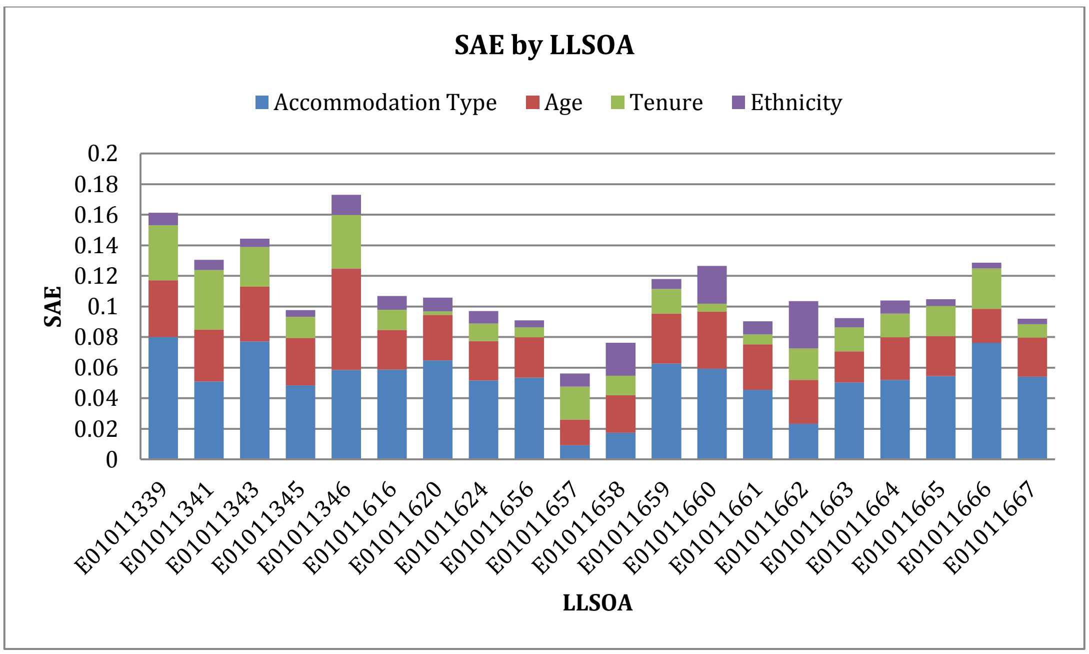


Figure 5.11 Illustrating the contribution of each validation variable to the total error using SAE by LLSOA

The limited variation in TAE values as reported in **Table 5.23** (TAE by rule set) can also be highlighted. Recall that the seven behavioural rules were chosen based on details documented in the qualitative literature. If these details are re-enacted as presented in the literature, then it is likely that a model designed to run with all seven behavioural rules should perform better than a model designed to run with only two or three rules. In other words, if households make decisions based on the *Ethnicity* and *Neighbourhood Quality* rules only as opposed to the full complement of the seven behavioural rules, then in theory the error statistics should be higher. Based on the output statistics reported, this is not initially apparent. However, as previously illustrated, if the TAE is observed at the LLSOA level then the disparity in results is more apparent. The graph below, **Figure 5.12**, is an example of this. Comparing the results from rule-set 61 (*Transport, Ethnicity, Social Status*) and rule-set 94 (*Ethnicity, Social Status, Schools, Known Areas*), at the OA level rule-set 61 consistently produces fewer errors than rule-set 94.

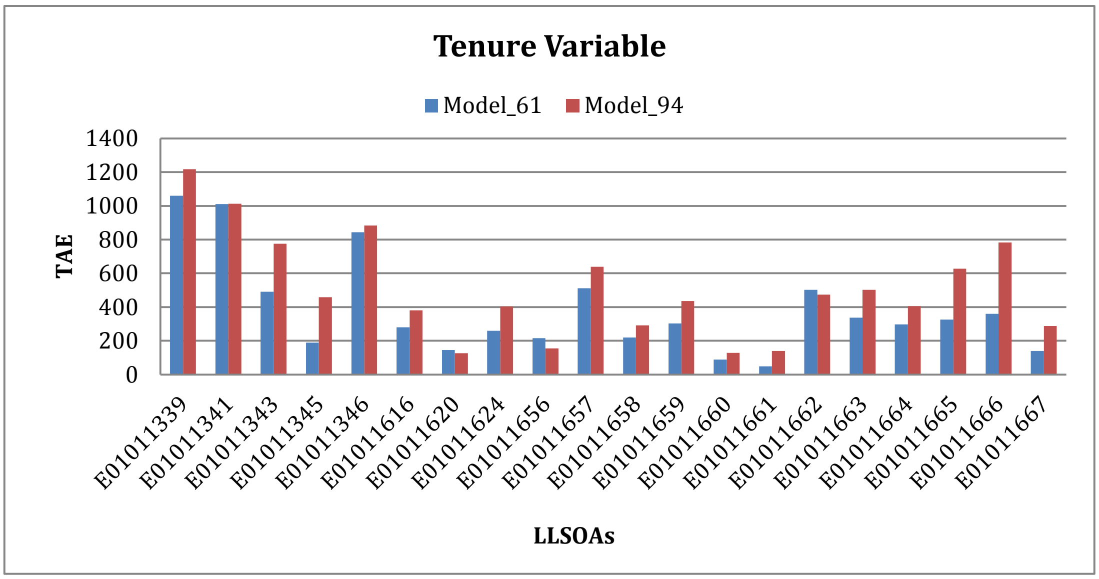


Figure 5.12 TAE at LLSOA level

Therefore, though at the EASEL level the TAE may appear similar, there is more variation in the actual distribution of households at the LLSOA level. Again, it is apparent that the aggregated results hide the variation at lower geographies therefore meaning that though the TAE and SAE figures may appear similar across rule-sets for the entire EASEL area, when these figures are disaggregated to the LLSOA geography, the variation is in error statistics is more apparent.

The lack of aggregate variation may also be linked to some of the assumptions made in the previous chapter (**Chapter 4**). The CHAIRS simulation model is presented as a closed system; there are no allowances made for households moving in and out of the EASEL district. Widening the geographic space in the system to include the entire city of Leeds or the county of Yorkshire may induce more variation. However, the boundaries must end somewhere. EASEL is chosen not only because it is the study area but also due to limiting computer capacity. With a limit of 48 hours maximum for one process on the available supercomputing environment larger simulations are currently impractical. Even if the execution time for a given model was one hour for ~35000 households, ~19 hours would be needed to run a model for the city of Leeds with a population of ~700,000. Given that at least 127 model runs are needed for the first calibration exercise alone, this amounts to a total of more than 100 days. Not only is this an exorbitant amount of time but it also does not include the second calibration phase or model re-runs. Such a computationally expensive procedure has not been possible under the current constraints of this project.

Also, the model does not currently include a direct mechanism for altering mortality, fertility or other family changes. Arguably these are the factors that drive the family life cycle; a process noted by Dieleman (2001) to be the most influential in residential mobility decisions. However, the model subsumes this information in the '*propensity to move*' statistic. The creation of a mechanism to drive the family life cycle would have been quite a large undertaking and will be considered for future work. If such a model is created, the behaviours which are encapsulated by the '*propensity to move*' statistics can be disaggregated and further explored.

5.3.3 Optimising the Best Rule-set Combination

In addition to examining the performance of varying rule-sets, the effects of altering other model parameters can also be explored in order to improve the goodness of fit to real-world data. Such parameters include distance measures in the transport rule and tolerance levels in the ethnicity rule. Technically, for each of the 127 rule-sets identified, the built-in parameters should be strategically altered and the results examined. A process such as this can be performed by the use of a genetic algorithm. Recall from an earlier section that genetic algorithms allow for automatic calibration (**Section 5.1**). For each rule-set, various parameter values can be automatically combined and the results tested to identify the combination producing the best fit to real-world data. Again there is a limitation on computing capacity, as if a genetic algorithm is used this automated process would need to be executed for each of the 127 rule-sets originally identified and is likely to take an inordinate period of time (**Section 5.2.1**). As a result of this, manual calibration of a single rule-set is used. Parameter values are chosen and are strategically altered during the model execution. The results are then compared. Though alternative parameterisation of sub-optimal rule-sets could improve their performance, for practical reasons only rule-set 61 is adjusted in this process. In addition to this, each parameter is varied separately though combinations of variations might improve performance. Such decisions were made taking into account the practical limitations of current computer processing levels.

Thus, using rule-set 61 (*Ethnicity*, *Transport*, *Social Status*), the parameters within the ethnicity and transport rules are systematically altered. The ethnicity rule originally states that at least 33% of surrounding neighbours in the new neighbourhood should be of the same ethnic group as the household wanting to move. This parameter is altered in 10% increments starting from 10% to a maximum of 100%. In addition to this, the transport rule as originally reported ensures that households live within one mile of a major transport route. This parameter is altered in one mile increments starting from one to a maximum of ten miles. First the tolerance parameter in the ethnicity rule is altered. **Table 5.26** illustrates the change in SAE as the tolerance level is changed.

Tolerance (%)	Ethnicity	Tenure	Age	Accommodation Type	Average SAE
10	0.260074	0.332766	0.61284	1.068366	0.590676
20	0.246386	0.330001	0.614715	1.066587	0.589568
30	0.253967	0.330899	0.627252	1.071689	0.599102
40	0.245697	0.325314	0.614651	1.073478	0.589718
50	0.244683	0.326898	0.615072	1.064401	0.588918
60	0.243066	0.3252	0.625189	1.070673	0.595611
70	0.247992	0.326835	0.611457	1.082058	0.589272
80	0.241419	0.317177	0.612586	1.073415	0.586867
90	0.228768	0.346833	0.613894	1.066043	0.588889
100	0.246267	0.321941	0.603324	1.068143	0.581622

Table 5.26 Change in Ethnic Tolerance by Average SAE

It is at a tolerance level of 100% that the average SAE is lowest. At this level, the age, and accommodation type variable perform better than at the 80/90% levels which also perform better, on average, than all other levels. It is at the 90% tolerance level, however, that the goodness of fit for the ethnicity variable is lowest, this is significant and favourable. Thus though the overall average at the 90% threshold is higher than the lowest average, this slight drop in total quality is felt worthwhile for the considerable gain in this key prediction area. Thus the 90% tolerance level is chosen.

The other parameter of interest is the distance parameter in the transport rule. **Table 5.27** below gives the results for this manual calibration exercise when the ethnic tolerance is set at 90% and the distance parameter in the transport rule is explored.

Distance	Ethnicity	Tenure	Age	Accommodation Type	Average SAE
1	0.622174	0.323014	0.223464	1.055162	0.555954
2	0.628198	0.348809	0.240807	1.031423	0.562309
3	0.632482	0.369974	0.229349	1.019069	0.562719
4	0.633212	0.366918	0.239827	1.02482	0.566194
5	0.634778	0.377492	0.238943	1.017432	0.567161
6	0.614801	0.371073	0.224648	1.025145	0.558917
7	0.635444	0.372637	0.236959	1.034742	0.569945
8	0.631877	0.372969	0.234883	1.033313	0.568261
9	0.620595	0.358861	0.237807	1.062061	0.569831
10	0.619051	0.330539	0.226035	1.08371	0.564834

Table 5.27 Change in Transport Distance by Average SAE

It is at a distance of one mile that the overall average SAE is most favourable. Here, the overall average SAE is regarded as most important as none of the validation variables directly capture the performance of the transport route distance parameter.

When these parameters are altered in this way, the performance of rule-set 61 is improved. The results are noted in **Table 5.28** below.

Model 61	Ethnicity	Tenure	Age	Accommodation Type	Average
[1]**Original**	0.251202	0.320578	0.616413	1.067922	0.564029
[2] **Ethnicity = 90% Transport = 1 mile**	0.223464	0.323014	0.622174	1.055162	0.555954

Table 5.28 Model 61 showing the results measured by SAE when additional parameters are manipulated

Where,

[1] Original = Ethnicity (33%), Socio Economic Status, Transport Routes (1 mile)

[2] Ethnicity (90%), Socioeconomic Status, Transport Routes (1 mile)

5.4 Critiquing the Performance of the CHAIRS Model

Thus far the methodology for calibration and validation has been discussed. Results from the CHAIRS simulation have been presented and the methodology applied. Rule-set 61 (*Ethnicity*, *Social Class*, *Transport Route*) was noted to be the best performing rule-set with ethnic tolerance set at 90% and transport distance set at 1 mile. Based on the results presented, the *Transport* rule appears to generate results in agreement with the literature. However, in the ethnicity rule, using a tolerance threshold of 90% generates the best fit when compared to the validation data set; this is interesting. Recall that the ethnicity rule mimics Schelling's theory loosely (**Section 4.3.3.2**). That is, unlike Schelling's rule which is concerned about with tolerance at the household's current residence, in the CHAIRS model, ethnic tolerance is taken into account in the neighbourhood of the new house. Thus by traversing the parameter space for ethnic tolerance, a more appropriate parameter can be gauged. Though 90% may appear high, it may be reflective of the high incidence of those of White ethnicity across the EASEL district as reported in the ROP validation dataset. Also, ethnic preference may need to be higher than Schelling found because segregation is high in this district in general. Note, however, that the other rules will work to mitigate the ethnicity effect. This may indicate that in limiting the number of variables to ethnic tolerance, Schelling counter-intuitively gained an over-optimistic opinion of the level of intolerance in segregated communities than was justified.

Key observations based on the technicalities of the model have also been presented. In this section, the results of the model will be compared to that in the literature, the section ends with a description of the spatial variances between the CHAIRS results and the ROP validation dataset.

5.4.1 Comparing the Qualitative Literature with the Quantitative Results

As mentioned, the rule-set chosen was comprised of the three rules; *Ethnicity*, *Social Status* and *Transport* routes. Using a tolerance level of 90% in the *Ethnicity* rule and a distance measure of one mile in the *Transport Routes* rule, this rule-set combination was identified as the best performing rule-set of all rule-sets and will be used to simulate behaviour when the scenarios are applied in the chapter to follow (**Chapter 6**). Using this rule-set combination with the additional alteration of the internal parameters, several observations can be made. Ethnic tolerance for diversity is very low, set at 90%. Though the EASEL district contains pockets of segregated minority groups, there is a large concentration of White households. Therefore, by increasing the tolerance value, this distribution appears to be effected more correctly. The distance measure in the *Transport Route* rule remains unchanged at one mile.

Compared to the literature presented in **Chapter 2**, the choice of these three rules sharply contrasts to the many factors noted to influence residential mobility behaviour. Though there are seven behavioural rules, only three rules have been used which produce the best fit to the validation dataset. Note that originally all rules were informed by the qualitative literature. Rule-sets containing behaviours based on social status, ethnicity and transport generate better results than rule-sets without such rules. To some extent this may mean that the interpretation of the qualitative literature could be refined so as to improve the general performance. Though this may help to improve the performance of the model, altering the model to fit the validation dataset is also not a useful solution as it may lead to over-fitting thus obscuring the actual dynamics of the model. In general this may be illustrating the difficulty in representing qualitative terms quantitatively.

On the other hand, the literature suggests that there are several factors to consider when residential mobility decisions are made. Nevertheless, the model is able to recreate a similar level of behaviour with a reduced number of behavioural rules. It may be the case that the qualitative literature may be overestimating the degree to which some rules play out the city-scale – this is, after all, one reason for this modelling exercise. Despite this, the rules are still able to recreate similar results as in the ROP data set. Also, due to the limitations of the ROP data, the fact that the CHAIRS results do not mirror the ROP data exactly is acceptable.

5.4.2 Examining the Spatial Distribution of Errors

The results of the CHAIRS simulation can be illustrated spatially in order to examine the actual distribution of errors. Here the results of the CHAIRS model are compared to results from the ROP micro data for the 2005/2006 time period. These results are presented for each of the four validation variables; ethnicity, tenure, accommodation type and age.

Ethnicity

The EASEL district, though widely populated by those classified as White British, is home to many ethnic minority groups. Communities are known to be segregated on this basis.

Figure 5.13 shows the distribution of Whites and Non-whites according to the ROP and the CHAIRS model. Analysing both ethnic groupings, there are parallels between the CHAIRS model and the ROP micro data. It is apparent that the CHAIRS simulation overestimates the White population in some areas while underestimating the non-White population in other areas. Notice, however, that the distribution of non-Whites are similar; that is, in both the CHAIRS simulation and the ROP dataset, high concentrations in one area are reported with the same high concentrations in both datasets though the exact population statistics may not be the same.

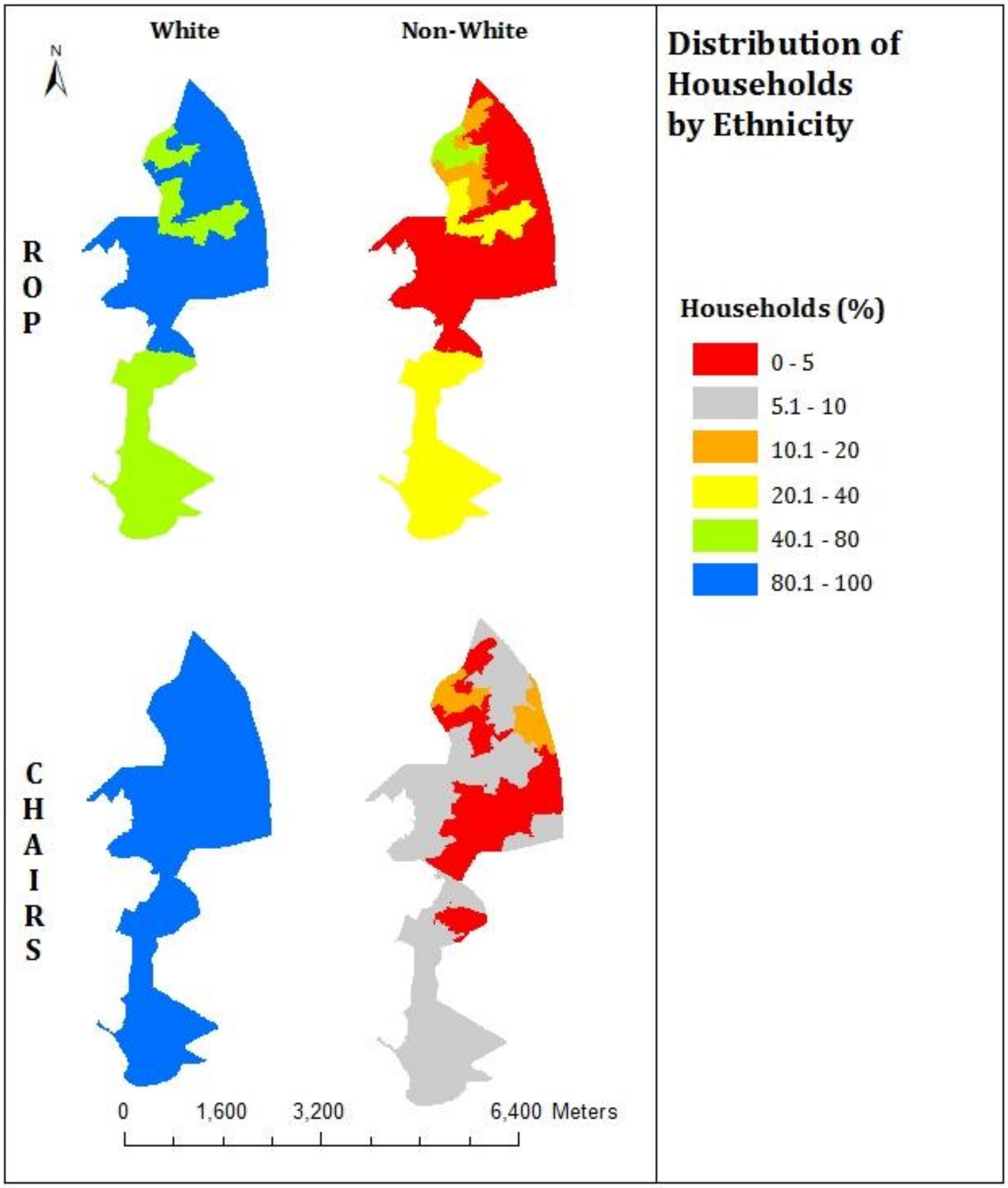


Figure 5.13 Spatial Distribution by Ethnicity

It can be seen that there is more spatial variation in the ROP micro data than that of the CHAIRS model, a phenomenon not uncommon in modelling (CROOKS, A *et al*., 2008). Such a phenomenon can be linked to the absence of precise details noted in the qualitative literature altering some of the dynamics of the bespoke model. Calibration is important to ensure that the real-world trends of interest are replicated. In this case, though the households distribution by ethnicity is not exactly the same in both datasets, there appears to be moderate ethnicity counts in the south of the illustrated district, low ethnicity counts in the centre and high ethnicity counts in the north when the non-white ethnic groups are compared in both the ROP and CHAIRS datasets. In general, there are high concentrations in the ROP throughout the district reported for the white ethnic group. Though these high concentrations are reflected in the CHAIRS simulation, the detail is reduced.

Also, noting that the ROP data is not perfect and that the CHAIRS model does not replicate significant events during the 2001-2006 periods, it is unlikely that the results of the CHAIRS model would precisely mirror the ROP data for the 2005/2006 time period. However, if the error statistics are observed, when aggregated to the EASEL area geography, the overall errors for the ethnicity variable are low indicating a favourable goodness of fit. Thus, the general distribution of households across the area appears to be reasonable using ethnicity as the test indicator. With this in mind, the other validation variables can be considered.

Tenure

Figure 5.14 below explores the tenure variable. As with ethnicity, though disparities exist, there are similarities that can be noted. Here some features of the ROP data are also reproduced by the CHAIRS model: private renting is highest in the south, concentrations of private renting are always low relative to social and owner occupier housing, ownership is highest in the northern area while social housing is least prevalent in the centre.

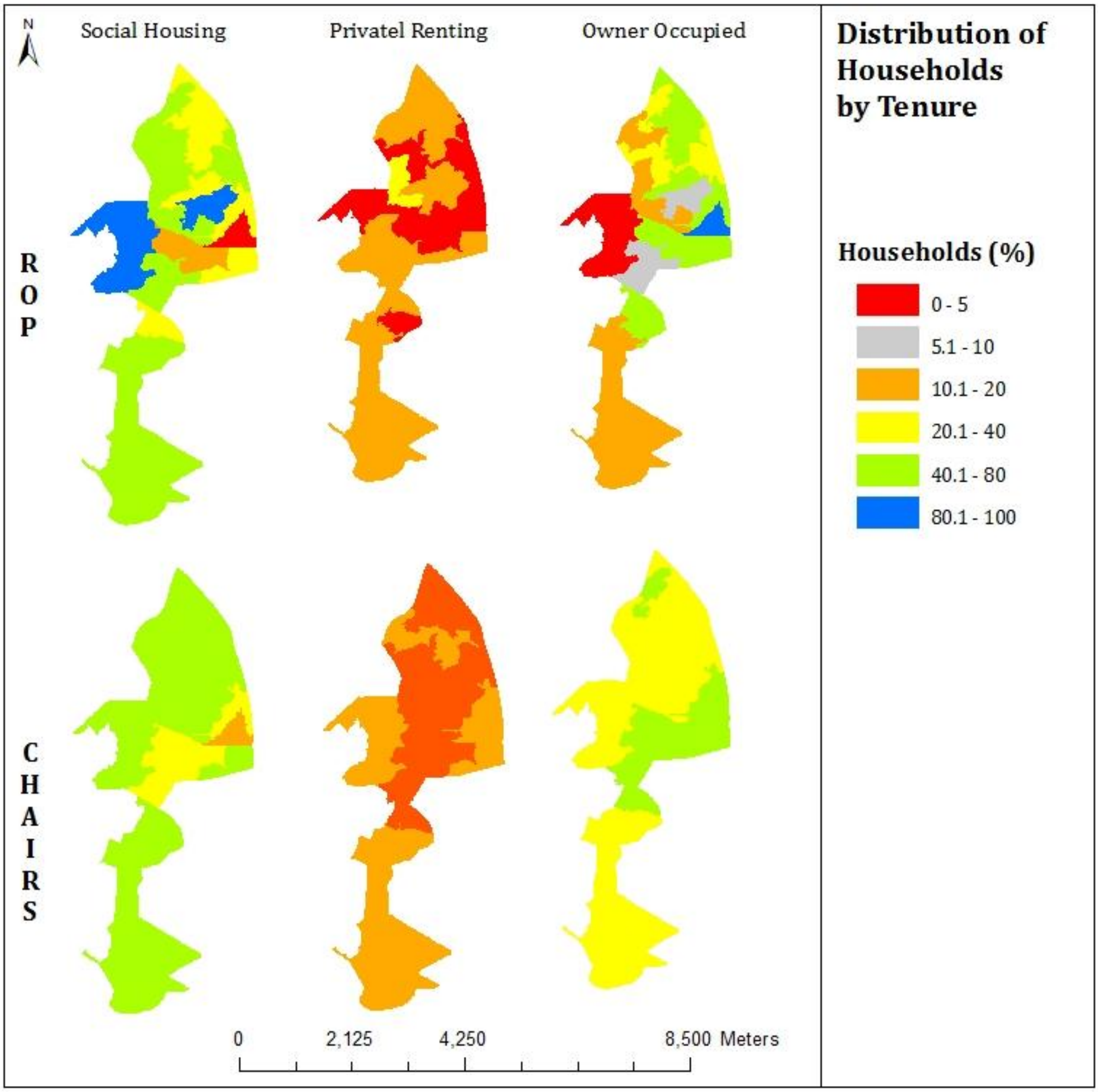


Figure 5.14 Spatial Distribution by Tenure type

Age

Unlike the ethnicity and tenure variables, the age variable produces higher error values when the TAE and the SAE are observed (**Section 5.3.2; Table 5.22** and **5.23**). As a consequence, not all trends observed in the ROP micro data are replicated by the CHAIRS model. **Figure 5.15** below highlights the differences and similarities between the ROP household distribution and the results from the CHAIRS simulation.

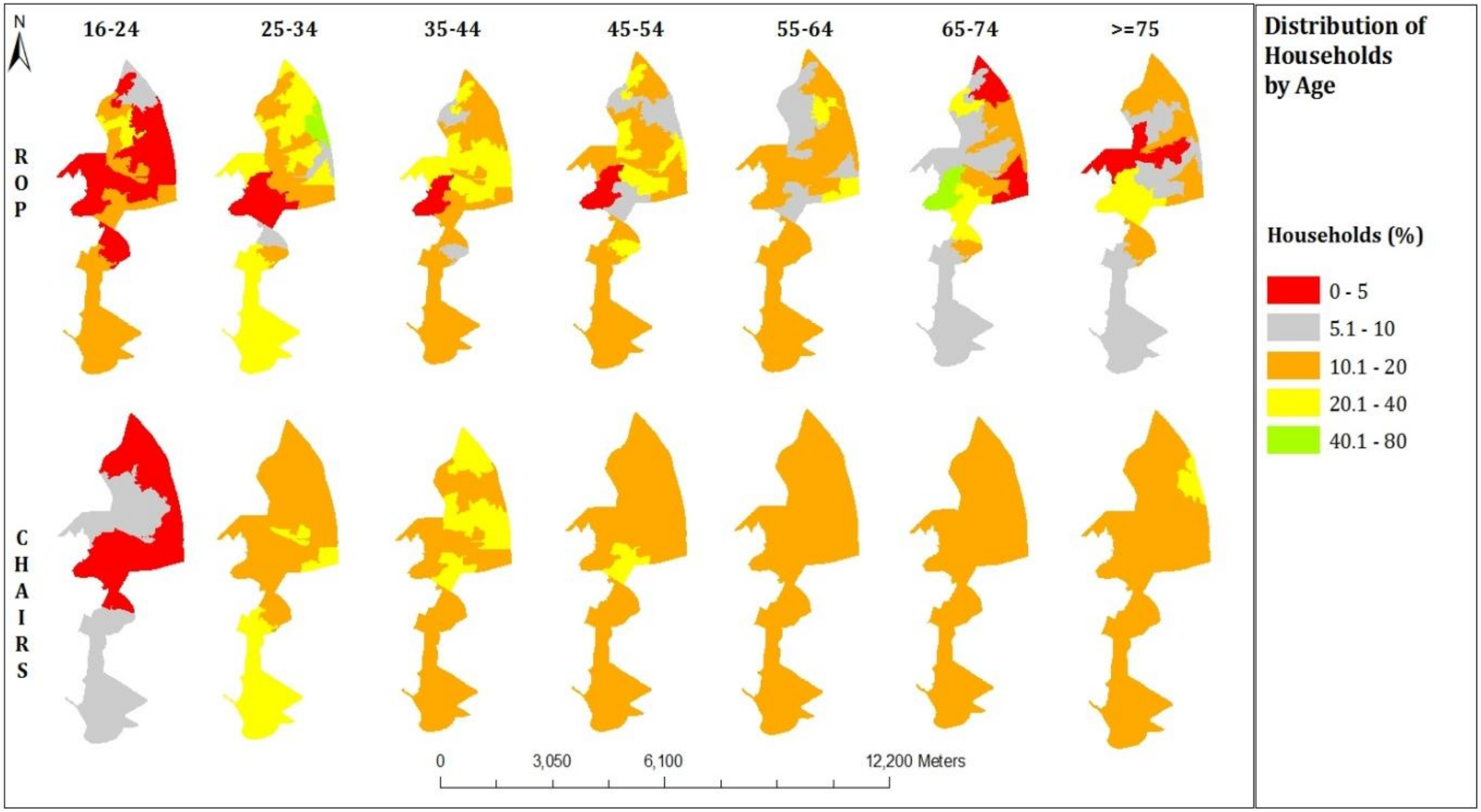


Figure 5.15 Spatial Distribution by Age

The map shows that there are stark similarities in the distribution of households in every age category. The 16-24 and 35-44 age categories all show similar trends. However, when the other categories are observed, household distribution trends are different. For example, in the 65-74 age category in the ROP dataset, there is a very high concentration of 65-74 year old household representative persons in at least one area. According to the CHAIRS model, there is a contrasting lower concentration of this same age category in the corresponding LLSOA. This may be because unlike the ethnicity and tenure variables, age has not been used explicitly in any of the relocation behavioural rules. This may account for why the error values reported for the age variable are not as low as those reported for the ethnicity and tenure variables. Age is encapsulated in the '*propensity to move*' statistic which is used to determine whether households wish to move or not (**Section 4.3.3.1**).

Accommodation Type

The accommodation type variable is the final one to be observed spatially. Much like the age variable, it was noted to be one of the validation variables which produced higher TAE and SAE values (**Section 5.2.3**). A comparison of the ROP household distribution versus the results of the CHAIRS simulation can be observed in **Figure 5.16** below.

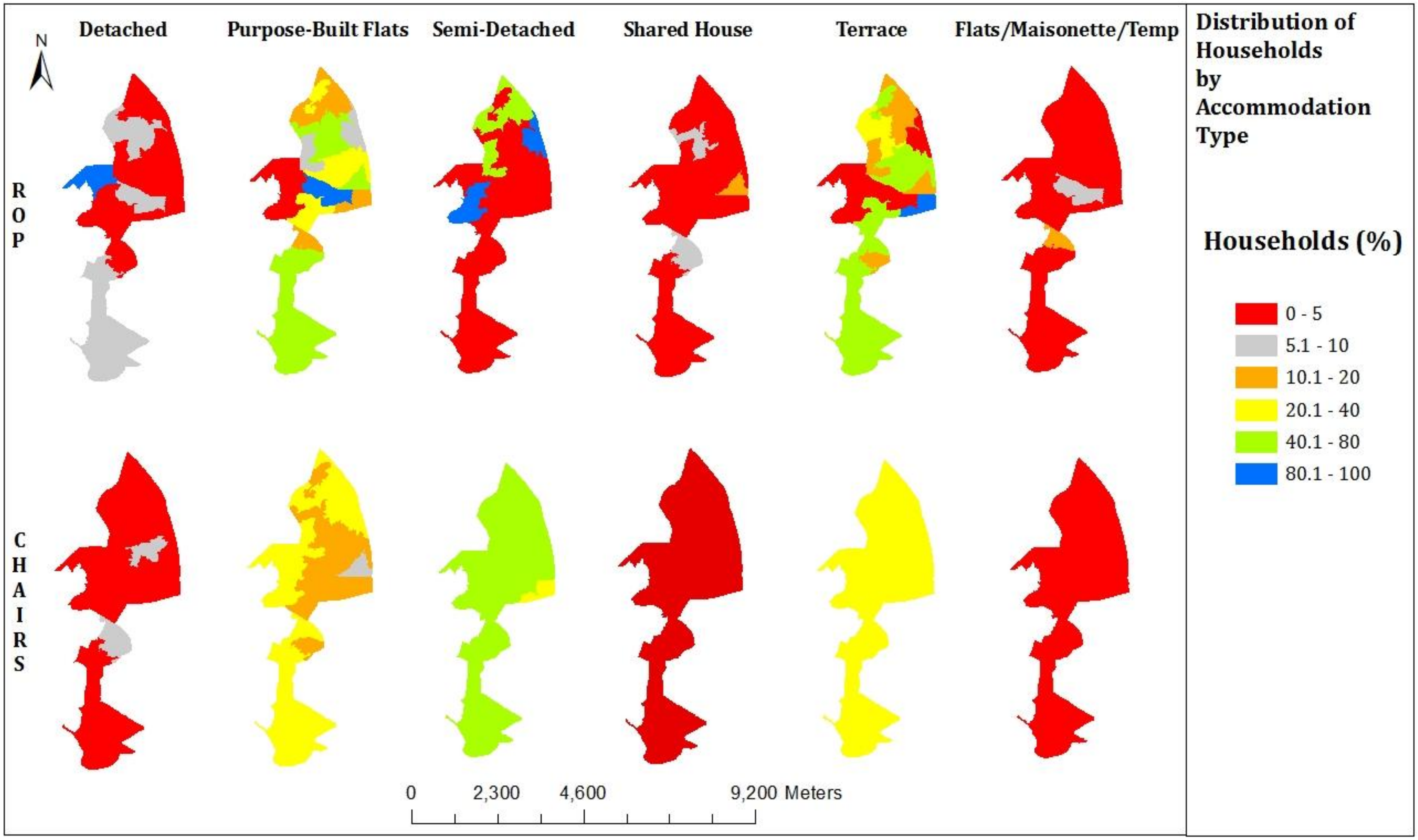


Figure 5.16 Spatial Distribution by Accommodation Type

The variances in this map are most apparent when the semi-detached and terrace housing categories are observed. The spatial distribution of households in detached, shared housing and flats/masionette/temporary accommodation are very similar according to the maps above while the CHAIRS simulation show less variation than the ROP when results for purpose-built flats are observed. In general, this indicates that some of the trends in the real world are being replicated by the simulation model.

Though these disparities may be linked to the fact that there is no relocation behavioural rule directly governing the type of accommodation chosen, it is important to note the assumption made when the rules were discussed (**Section 4.5**). The model also assumes that the physical distribution of houses is the same through the 2001-2006 period meaning that changes in the physical distribution of housing remains unchanged throughout the lifetime of the model. The original 2001 polygon data for houses was obtained through the Ordnance Survey. At the time of modelling, more recent data could not be obtained and it was thought that changes of note did not take place over this time period (GJESSING, M., 2010).

The general conclusion here is that though there are similarities in the spatial distribution of households when the four validation variables, they exist some differences when the ROP micro data is compared to the results from the CHAIRS model for the 2005/2006 period. The ROP dataset is a powerful source of information but new to academicians thus it may be the case that more experience is needed to understand the spatial and compositional biases in this dataset. Similarly, with ABM, it is difficult to determine the exact cause of these differences. On one hand, some details could have been lost when translating information from the qualitative literature for suitable use in this quantitative model thus showing the difficulty of coverting qualitative understandings into definitive quantitative terms. On the other hand, the assumption was made that the '*propensity to move*' statistic is sufficent to recreate the residential choice decison. The CHAIRS model does not include mechanisms for altering the ages of households or mechanisms to generate births and deaths. As a result of this, the households population used at the beginning of the model in 2001 is used throughout the model without change and is also linked to the overall assumption that though the population may change over time, the EASEL area is sufficiently diverse meaning that the same type of people are likely to be present over the lifetime of the model. In restrospect, it may have been useful to implement a mechanism to alter the population demography.

Amidst these assumptions, the trends observed in the ROP micro dataset are reasonably similar to the spatial variations observed in the CHAIRS model for the 2005/2006 period. This is particularly true when the fit for ethnicity and tenure are observed. It may be possible to include additional rules and/or parameters to improve the performance of the model, though this does not gaurantee producing better results.

Overall, the error statistics have been analysed with a critique on the spatial distribution of households presented. Suggestions as to cause of errors and variances in the spatial distribution between the ROP and the results of the CHAIRS simulation have been noted. This has been linked to the assumptions made in the methodology of the model as well as general difficulties related to ABM (**Section 4.5, Section 3.2.4** respectively). What should be highlighted is the fact that modelling is not an exact science and models are abstractions of reality. Therefore care must be taken when presenting any results that attempt to forecast housing distributions in the next chapter.

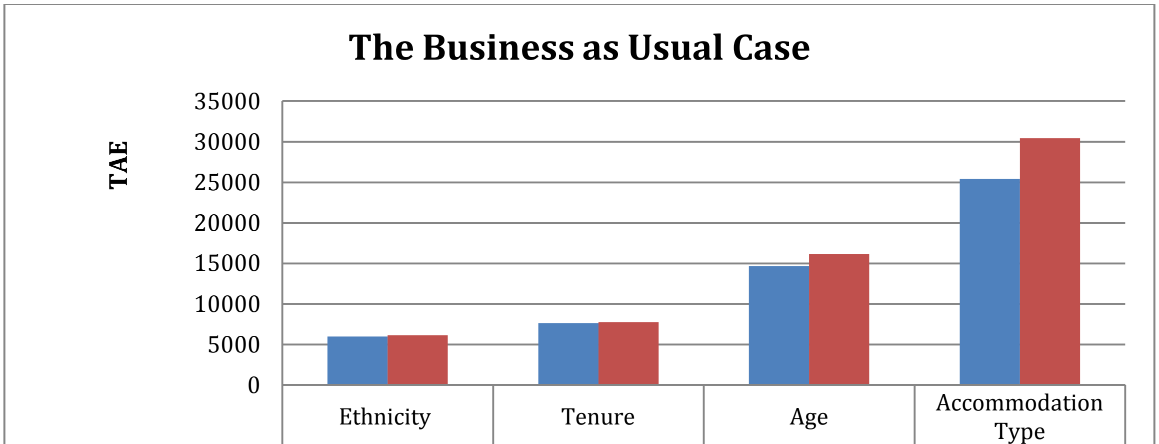
5.4.3 Benchmarking

Now that the optimisation procedure has been discussed in detail, it is useful to draw conclusions about the general performance of simulation model. Noting the difference in results when the model is compared to the ROP validation dataset and the differences in some of the spatial patterns when analysed, where does the CHAIRS model stand? As earlier mentioned, in simulations such as these, there will always be differences between the validation dataset and the results of the model, however, it is the level of acceptible error that may be in question. In this model the error ranges from ~11% in the tenure variable to ~50% when age is observed. One would imagine that an error level of 11% may appear acceptable for a population of over 35000 households while the converse may be true for 50% of error. However, many of the trends observed in the CHAIRS results are similar to the ROP validation dataset when observed spatially.

Having said this, it may seem that the CHAIRS results could be better. A likely expectation may be that the absolute error differences should be closer to zero and the trends observed in the ROP validation data should be mirrored by the CHAIRS simulation. Without a similar agent-based model to compare these results to, it is difficult to label the model as performing poorly. It may be possible that altering the present design would yield better results, for example changing the design of the behavioural rules. Conversely, it is possible that such alterations would yield results similar to those produced by the CHAIRS model. What is known is that the model replicates aspects of reality to a reasonable degree and this is evident when the spatial data is analysed.

In the absence of a model of comparable functionality, the performance of the model at the 2005/2006 period can be compared to a 'business as usual' case, taken as the inactive case, to the performance of the model at 2001 when no rules were applied. Such a comparison indirectly tests the effectiveness of the behavioural rules. Here the 2001 census data is taken as the 'no change' model results and validated again the 2005/2006 ROP micro data to generate an error statistic. This is compared with the error statistic generated by comparing the proper model 2005/2006 simulation results with the 2005-2006 ROP micro data. The model producing the least number of errors is thought to be the model which is a better predictor of the population distribution in 2005-2006. In general it is hoped that a dynamic model will generate better predictions than doing nothing.

In **Figure 5.17** below, the 2005/2006 CHAIRS results are compared to the 2001 Census results. Both sets of results have been compared to the ROP validation dataset and the TAE recorded for each of the validation variables. Using these validation variables, the 2005/2006 CHAIRS results produce consistently lower errors when compared to the validation dataset than the 2001 Census. Much like earlier observations, the ethnicity and tenure variables produce a better fit to the ROP validation dataset. Also, when the age and accommodation type variables are observed, the difference in error is noted to be much higher than when the ethnicity and tenure variables are observed. Despite this, the results suggest that the behavioural rules in the CHAIRS simulation model have made a difference when rule-set 61 is used; the CHAIRS simulation produces a somewhat better fit to the validation dataset than the 2001 Census and is plainly replicating some of the system dynamics across the modelled years.



	Ethnicity	Tenure	Age	Accommodation Type
2005-2006 CHAIRS	5977	7627	14666	25408
2001 Census	6135	7769	16187	30426

Figure 5.17 The Business as Usual Case using the TAE

Note that these results capture all error at the EASEL area level. Examining the results at a lower geography, the differences between the CHAIRS results and the national Census are more apparent. **Figure 5.18** below highlights this; the CHAIRS simulation is shown to provide a better match to the ROP validation data than the census when the semi-detached category of accommodation type is analysed. Here, the CHAIRS model produces consistently lower results than the 2001 Census. A similar pattern of difference can be observed for the other validation variables used.

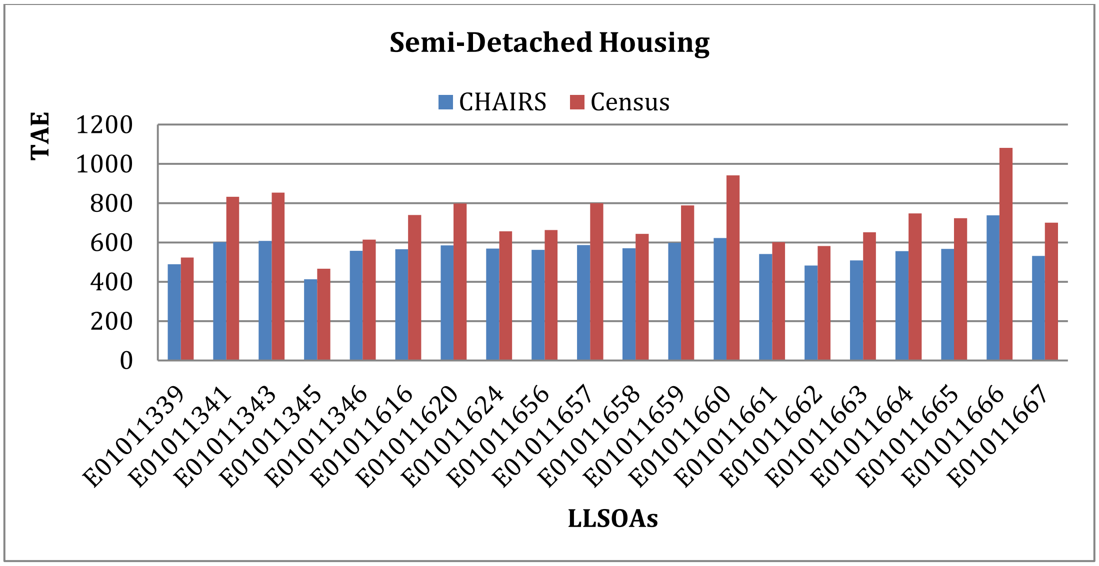


Figure 5.18 The Business as Usual Case using the TAE at the LLSOA Level for Accommodation Type

As presented in the description of this project (**Chapter 1**), the aim of the project is to assess the potential effects of regeneration schemes in the EASEL district. Regeneration schemes have been earmarked for the area based on the premise that the area suffers from zones of concentrated poverty and segregation. If the model is able to create segregation patterns then this is another way to illustrate its quality. By using the Diversity Index (DI), the level of socioeconomic segregation can be assessed. The DI for an area measures the likelihood that two randomly selected people would differ based on some predetermined factor (BREWER, C. and Suchan, T., 2001). In this case, this difference would be measured using ethnicity and the socioeconomic group. High values mean that the area in question is more diverse while low values indicate less diversity. The index is defined as follows, where **n** is the total number of ethnic/socioeconomic groups, **P** is the proportion of households in area **i**:

$$DI_i = 1 - \sum_{s=1}^{n} \left(\frac{P_i^s}{P_i^*} \right)^2$$

In general, the CHAIRS results predict a low level of ethnic diversity in the EASEL district. Between the 2001-2006 period, the modelled ethnic diversity ranged from 10% to 20% when observed at the LLSOA level. On the other hand, the modelled socioeconomic diversity is reported to be higher ranging from a low of 35% to a high of

~85%. On first examination, the aggregated CHAIRS results appear to show that the EASEL district becomes more diverse over time when the socioeconomic status variable is observed at the LLSOA level. However, when analysing the model data over a longer time period and by drilling down to the OA level, out of the 287 OAs making up the EASEL district there are 138 which become less diversed in the 2001-2006 period. This is shown in **Figure 5.19** below. Notice the peaks and troughs in diversity record, it seems apparent that the diversity of this OA is reduced over the 20 year simulation period. These are the type of trends which are seen in reality.

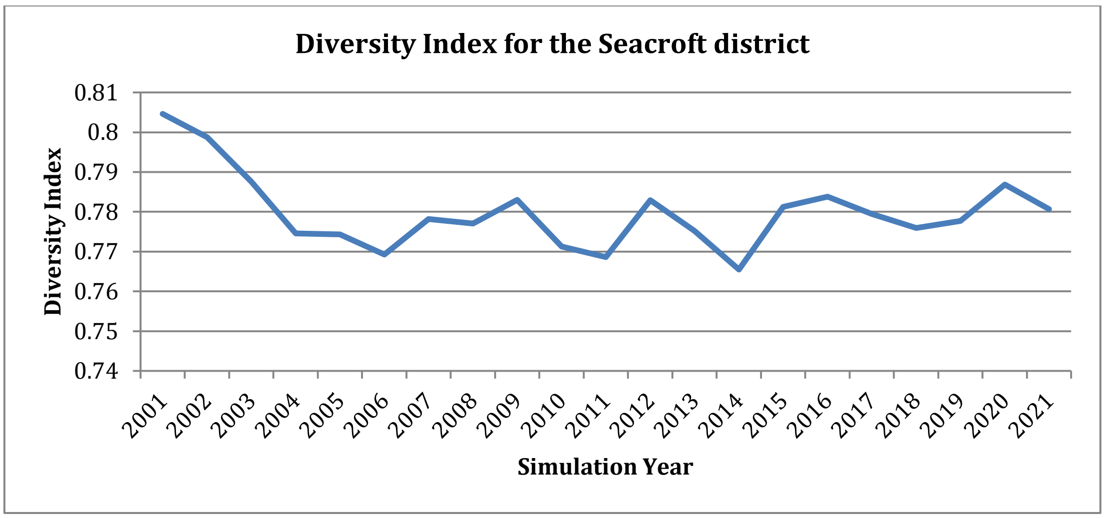


Figure 5.19 DI by Social Class for Seacroft district

Based on what is known about the EASEL district, there are areas of low ethnic and socioeconomic diversity. Though the lack of ethnic diversity may not be damaging, it is the combination of low ethnic and socioeconomic diversity that may be problematic. Again, this trend is replicated by the model and these are the types of trends which help to instill some confidence in the model as it is. Further analysis of trends such as these will continue in the chapter to follow (**Chapter 6**). Finally, due to the constant movement of households initiated by the predefined 14% movement rate per year, the model does not reach a state of equilibrium where residential mobility behaviour is reduced or halted. This is also linked to the closed nature of the model where demographic details do not change throughout the lifetime of the model. For this reason, simulating dynamic equilibrium is a challenge. In reality, population dynamics do change; ethnic populations have been growing and will continue to grow

(WOHLAND, P *et al*., 2010), however, this trend is not fully encapsulated in this model. As a consequence, assessing turnover is not possible using this framework. This was previously outlined in the aims and objectives (**Section 1.2**).

5.5 Conclusion

The chapter detailed the methodology for calibrating and validation the CHAIRS simulation model. Results from these processes have been presented with analysis on the variances between the qualitative and quantitative literature discussed. Spatial similarities between the ROP validation dataset and the CHAIRS simulation results were also highlighted. Overall, some variables in the model perform better than others with the best performing variable reaching an error rate of ~11% at the EASEL area level and 0.2% at the LLSOA level. Though it is difficult to set precise benchmarks for this model when compared to others models, what is known is that trends similar to those observed in the ROP dataset are replicated and segregation trends are similar. Based on these conclusions, the model will be used to observe the trends over time when scenarios are applied in the results chapter to follow (**Chapter 6**).

Chapter 6
Exploring the Policy Implications Using 'What If' Scenarios

6.1 Introduction

The CHAIRS model has been created to simulate housing choice behaviour in order to assess the likely effects of regeneration policy on the EASEL district of Leeds. In this chapter, the results of two of the proposed regeneration projects will be presented and discussed.

Recall from **Chapters 1** and **2** the background of the EASEL regeneration scheme. The EASEL district is known to be a densely populated neighbourhood of approximately 35000 households. The area is home to a diverse set of ethnic minority groups concentrated in a few zones. It is also an area which is home to many households who rely on social housing provisions, with higher than average indices of deprivation recorded for some of the EASEL districts. There are high incidences of crime and unemployment, and problems such as these reduce the sustainability of neighbourhoods as they are not able to attract businesses and, by extension, provide employment opportunities (**Section 2.4.2**). In general, quality of life may be negatively affected for households living in communities such as these.

As a result of this, Leeds City Council has proposed a range of regeneration plans which should bring improvements to the area. One such plan is the building of new mixed tenure housing developments. Here, households are mixed by socioeconomic status in an arbitrary pepper-potting style where houses are all very similar physically and residents are oblivious to the other housing tenures in the development. The policy of using mixed tenure communities to fuel change is founded on the notion that it will diffuse pockets of poverty by redistributing those reliant on welfare support throughout a community with others not reliant on welfare support. As discussed in

Section 2.4.3, by using a role model approach, more progressive households among mixed communities may be able to influence other households and over time effect change. For example, those unemployed and able to work may be encouraged to seek employment as opposed to relying on welfare support.

Another project falling under the remit of regeneration is the proposed change in the road network connecting the north and south of the EASEL district. Though the road network already exists in parts – it is a series of minor roads – such a change would require road widening in order to transform the present network into a major road. An improvement such as this would transform this route into a major transport artery and could help to improve access to work for those employed in the city centre; bearing in mind that there is a relationship through car ownership between being employed and seeking housing within reasonable distances of major transport routes. A third regeneration scheme is added by implementing both schemes simultaneously. The CHAIRS model is adapted in such a way so as to simulate the implementation of the three regeneration schemes.

By implementing the regeneration projects in the CHAIRS simulation, a series of 'what if' questions can be asked and answered. What if a new mixed tenure development is built? What if the road network is changed? What if a new mixed tenure development is built AND the road network is changed? What are the outcomes from these policy interventions in terms of the impacts on the distribution of the population? Are the results socially acceptable? In order to assess this, the three scenarios will be compared to the results of the CHAIRS simulation when none of the regeneration schemes are implemented: this will be referred to as the baseline situation. Note that this is an exercise in forecasting and as such all results generated are likely to give broad based ideas of what could possibly happen if the regeneration schemes are implemented in reality.

Each scenario is implemented using rule-set 61 as identified in the calibration/validation process of the previous chapter (**Chapter 5**). This rule-set combines the *Ethnicity*, *Socioeconomic Status* and *Transport* rules. For the baseline situation, the CHAIRS model is run from 2001 until 2021 using rule-set 61. There are no policy interventions at the baseline and as such may be likened to the worse-case outcomes for the EASEL district. For scenario 1, the shapefile used to represent houses is amended to include the new mixed tenure development. For scenario 2, the shapefile used to represent the road network is altered so that the new road network is represented as a major road. Scenario 3 is a combination of scenarios 1 and 2.

All scenarios are implemented in 2001 thus the outcomes can be observed from year 2002 onwards. The scenarios are run for a 20 year period and end in 2021. This simulation period is the same for the baseline situation. An alternative to this would be to implement the scenarios in 2007 or another chosen year between 2001 and 2021. It should be noted that though this is not the methodology adopted, the results generated are no less valuable when implementing the scenarios in 2001; any changes observed will be discussed over the time period during which the scenarios have been implemented.

Results are analysed comparatively from one year to another using years 2011 and 2021 as strategic points as they represent the today situation as well as 10 years in the future respectively. Data is analysed at the OA level, of which there are 286 OAs. Due to the large number of OAs under observation, for scenario 1 the Gipton new development area will be discussed in detail while for scenario 2 the OAs surrounding the new road network will be discussed in detail. If other results arise that deviate significantly from the baseline situation, these will also be discussed. Note that OAs are the smallest census geography available and contain approximately 125 households. In order to analyse changes between the baseline situation and the scenarios, the change in the spatial distribution of the population is analysed using the DI, with socioeconomic status being the household attribute for consideration. The index of segregation (IoS) is also used to analyse the distribution of the population by ethnicity.

Recall from the previous chapter that the DI for an area measures the likelihood that two randomly selected people would differ based on some predetermined factor (**Section 5.4.3**). The index is defined as follows, where n is the total number of ethnic/socioeconomic groups, P is the proportion of households by ethnicity type or social class in area i:

$$DI_i = 1 - \sum_{s=1}^{n} \left(\frac{P_i^s}{P_i^*}\right)^2$$

The IoS measures the geographical distribution of two population groups. More practically, it represents the proportion of one population group who would have to move in order to be distributed in the same way as the rest of the population (Brewer and Suchan, 2001). The index is defined as follows, where b is the base population, r represents the rest of the population and P is the proportion of the households in area i:

$$IoS(b \ and \ r) = 0.5 \sum_{b,r=1}^{n} \left| \left(\frac{P_i^b}{P_*^b}\right) - \left(\frac{P_i^r}{P_*^r}\right) \right|$$

In addition to these indices, the Mann-Whitney U test and the Kruskal-Wallis test are used to test whether the results of the scenarios are comparably different from the baseline situation. The Kruskal-Wallis test is an extension of the Mann-Whitney test and is used for larger dataset. Both methods test the null hypothesis which assumes that the datasets under examination are the independent of each other ((MANN, H. B. and Whitney, D. R., 1947); (TRIOLA, M. F. *et al.*, 2007)). Both statistical tests are generated in SPSS which reports the whether the null hypothesis should or should not be retained for the datasets under examination.

Graphs and maps will be used to analyse the change in the spatial distribution of the household population over the 20 year simulation period. What follows is a presentation of the actual simulation results. The results will be analysed first using the baseline situation; this is followed by the presentation of results for each of the three scenarios. Comparisons are made between the scenario results and the baseline. The chapter concludes with a discussion on the broader themes that can be drawn from the analysis of the 'what if' scenarios with particular reference to general criticisms of urban regeneration, namely issues such as gentrification.

Note that other possibilities such as school impact might be considered in order to assess turnover rates as indicated in the aims and objectives (**Section 1.2**).

6.2 The Baseline Situation

Using the baseline case, the results of the model will be analysed to illustrate the possible housing choice decisions if there are no policies implemented over the 2001/2021 period. The table below (**Table 6.1**) is a summary of the movement counts throughout the 20 year baseline simulation. The results of this table are very similar from scenario to scenario, as the type of households that move is driven by the decision tree defined in **Section 4.3.3.1**. Here, at least 45% of the total population has moved over the life-time of the model. With a movement rate of 14% each year (**Section 4.3.3.2**), there are over 100,000 moves over the 20 years. On average, households move six times though the number of moves per mover household ranges from a low of 1 to a high of 13.

Description	**Count**
Total number of households	35729
Total number of households who moved	16387
Total number of moves	100038 (35729 * 0.14)* 20 years
Average number of moves	6
Highest number of moves by one household	13
Lowest number of moves by one household	1

Table 6.1 Summary of Total Movement Counts

Recall that the DI is used to measure the diversity of households in the EASEL district. In this case, socioeconomic diversity is measured. In general, the DI over the 20 year period fluctuates for all OAs. The index ranges from a low of 0.369476 to a high of 0.8752. In real terms this means that OAs are noted to be low in socioeconomic diversity at 36%, through to a reasonably high rate of diversity at 87%. There are four

OAs which dip below the 50% diversity mark and another five OAs which dip below the 60% diversity mark. Throughout the baseline model run, the diversity increases over the 20 year time period in each of these areas.

These nine areas are listed in **Table 6.2** and correspond to districts in Seacroft, Gipton, Harehills and Richmond Hill. Indices of diversity and segregation are noted in the table. Social class is used to determine the DI while ethnicity is used to determine the IoS. As the model is rolled forward, each OA noted increases in diversity staying within the 40-60% diversity range.

LLSOA Name	LLSOA	Output Area	Diversity Index at OA	Index of Segregation at LLSOA
Seacroft	E01011661	00DAGE0048	0.37	0.17
Harehills	E01011675	00DAGF0069	0.45	0.09
Gipton	E01011346	00DAFF0047	0.45	0.15
Seacroft	E01011658	00DAGE0011	0.49	0.09
Harehills	E01011679	00DAGF0066	0.50	0.14
Richmond Hill	E01011626	00DAGB0045	0.51	0.13
Richmond Hill	E01011619	00DAGB0015	0.52	0.2
Gipton	E01011346	00DAFF0023	0.54	0.15
Seacroft	E01011658	00DAGE0012	0.60	0.09
Gipton new development	E01011431	00DAFM0025	0.80	0.13

Table 6.2 Segregation Indices for areas of lowest diversity in 2001

Though there are areas of very low diversity as illustrated in **Table 6.2**, overall there is still a reasonably high level of socioeconomic diversity computed over the entire EASEL district. During the 20 year CHAIRS simulation run, the district has an average DI of 79% with very few areas falling below the 50% diversity mark. The converse is true for ethnic segregation; households in the EASEL district are predominantly White and as a consequence, ethnic segregation is noted to be very high. On average, segregation is recorded at 22% using Whites and Non-Whites to categorise the ethnic groups.

These observations bear a striking resemblance to the diversity situation described by the Leeds City Council (GJESSING, M., 2010); below average socioeconomic diversity in some EASEL district communities. This is coupled with very high incidences of ethnic segregation in the same areas, a situation thought to encourage concentrations of poverty and that forms the basis for the regeneration plans envisioned for the area. Areas such as these are illustrated in **Table 6.2** above.

Examining the diversity statistics more specifically, out of a total of 286 OAs, there are 138 OAs which have lower diversity indices in 2021 than in 2001 in the baseline situation. One example of this is illustrated in the graph below (**Figure 6.1**) for one OA in the Seacroft ward. Here, the DI was noted to be over 80% in 2001. Over the 20 year simulation period, however, this DI dipped to ~76%, reaching a level of 78% by 2021. The change in diversity is ~2% between 2001 and 2021. Though not a large change, it illustrates how the model tends toward segregation when the baseline results are analysed.

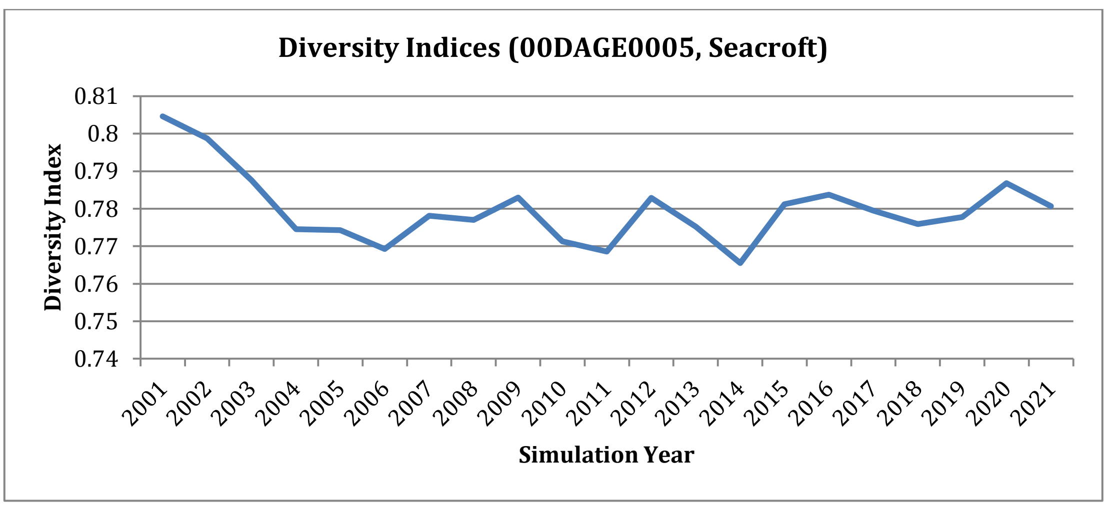


Figure 6.1 Diversity Indices for OA 00DAGE0005, Seacroft for the 20 year simulation period

Also in **Figure 6.2** below, the DI is considered over a range of three OAs of contrasting diversity. Here diversity ranges from very high values to very low values. In this illustration, OA 00DAGE0018 remains consistently high indicating that this OA is very diverse when the social class variable is observed. On the contrary, this is not true for the remaining OAs used in this example. Here, the diversity indices are noted to be considerably less. Both OAs 00DAGE0018 and 00DAGE0048 can be found in Seacroft while OA 00DAGF0069 is located in Harehills.

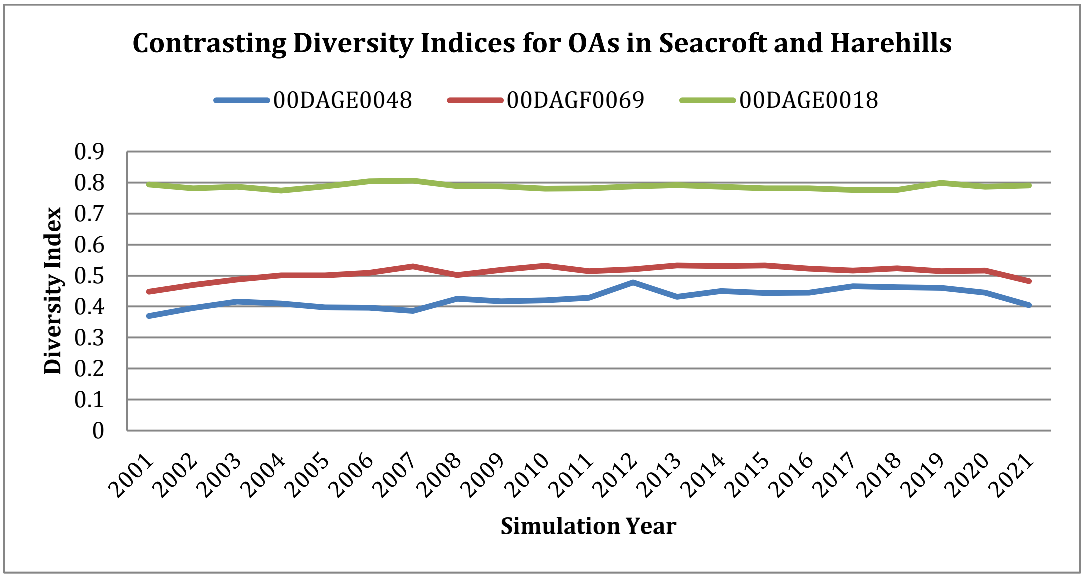


Figure 6.2 Diversity Indices for a range of OAs in Seacroft and Harehills over the 20 year simulation period

Scenario 1 will create additional housing in a mixed tenure development. This new development is to be built in the Gipton area of Harehills. **Figure 6.3** below illustrates how the DI changes in this area without the new mixed tenure housing development. Over the 20 year simulation period, the area sees a reduction in the level of socioeconomic diversity. This trend suggests that without policy intervention, this area will tend closer and closer to segregation. There appears to be a sizeable dip in the index between 2011 and 2012. This is indicative of a large number of low income households entering the area in 2012. The area is one which already has a higher level of low income households and as such this increase further reduced the level of diversity.

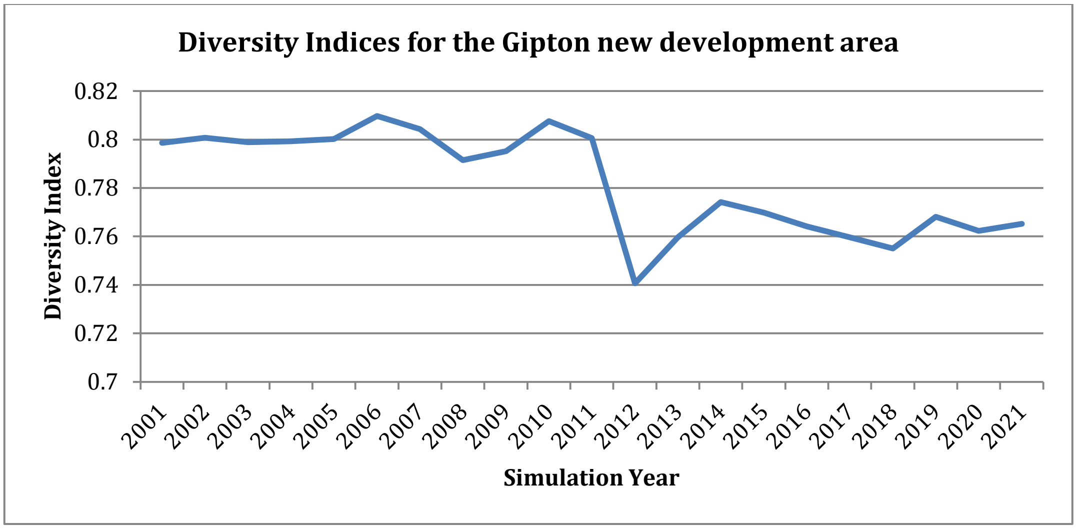


Figure 6.3 Diveristy Indices for the Gipton new development area at the baseline over the 20 year simulation period

When the socioeconomic categories are examined more closely, **Table 6.3** shows an increased amount of households in most socioeconomic groups for this area. Here the '*Intermediate Occupations*' and '*Small Employers*' categories decrease by 3-4% over the 20 year simulation period. The continued growth of those classified as '*Not Classified*' point to the low income character of this OA.

	Percentages		
	2001	**2011**	**2021**
Higher Managerial	0	0.8	0.8
Lower Managerial	6.4	9.1	8.8
Intermediate Occupations	6.4	4.1	3.2
Small Employers	7.9	3.3	2.4
Lower Supervisory	7.1	6.6	7.2
Semi-Routine Occupations	15.9	15.7	11.2
Routine Occupations	14.3	15.7	16.8
Never Worked, Long Term Ill	5.6	9.1	8
Not Classified	36.5	35.5	41.6

Table 6.3 Distribution of households in Gipton by socioeconomic status at the baseline

Scenario 2 allows for a change in the road network. In order to analyse the resultant change in household behaviour, the OAs surrounding this new road network are examined. **Figure 6.4** shows the diversity indices for year 2001, 2011 and 2021. The results below show some variation in the diversity levels over the simulation period. In several OAs, more diversity is reported over the 20 year simulation period, however in general there is little change.

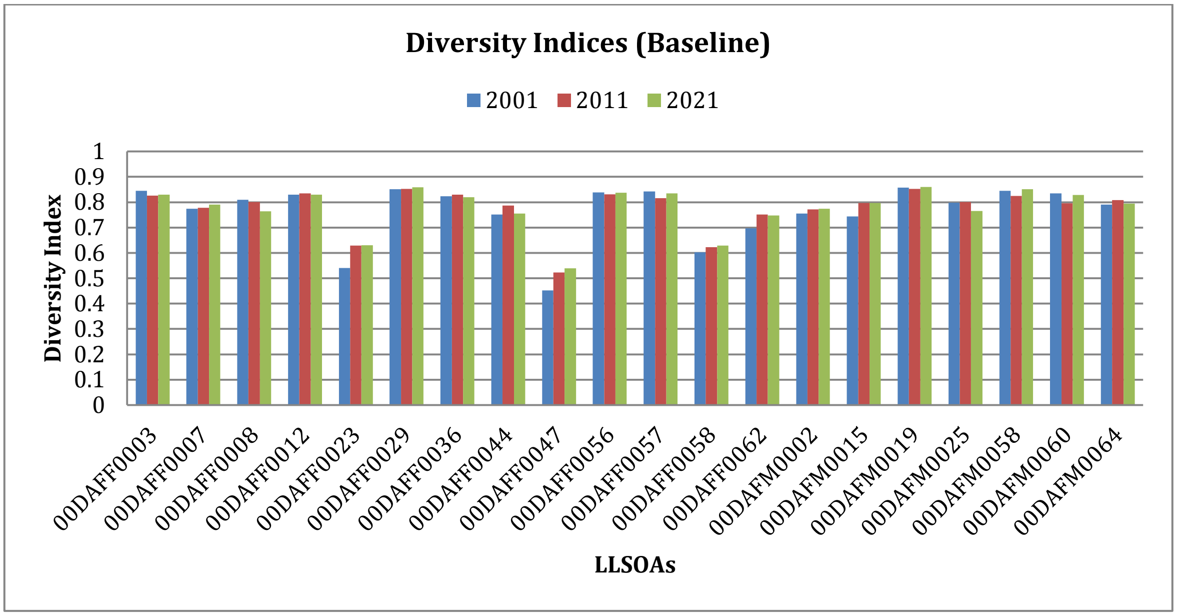


Figure 6.4 Diversity indices for districts surrounding the new road network at the baseline

Examining these results more closely, **Figure 6.5** support these claims when the change in population counts are observed. There are noticeable increases in the population counts in some areas while the reverse of this is true for other areas. However, the increase in the population size for the affected OAs appears to be small.

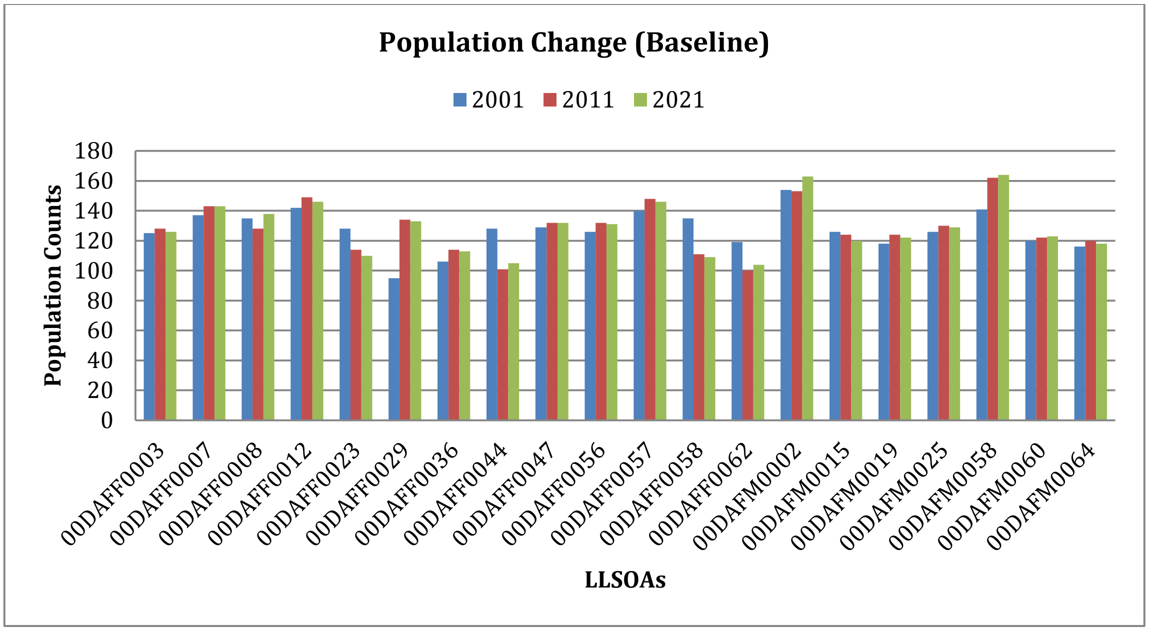


Figure 6.5 Population counts for districts surrounding new road network at the baseline

Maps can also be used to examine the change in population distribution during the simulation period. It should be noted that the OAs in the southern regions of the map are disproportionately large due to the large green corridor in this area of Leeds. Though these areas contain on average 125 households, when shaded on the map this may be misleading and as such caution should be exercised when viewing all map results. **Figure 6.6** has been used to illustrate this.

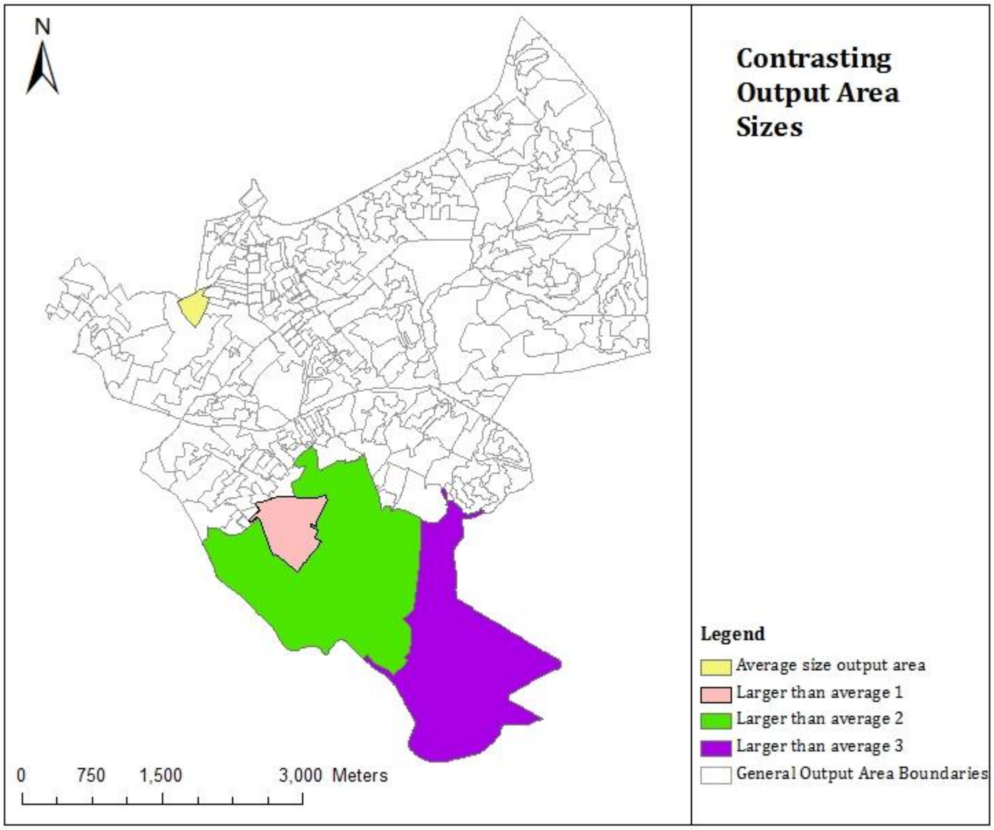


Figure 6.6 A Map Illustrating the Contrasting OA sizes in the EASEL district

The population distributions for the baseline situation across the 20 years are illustrated in **Figure 6.7** below. Larger pockets of concentration are signified using darker shadings. The darker pockets of '*Higher Managerial*' careers tends to disappear over time, so there seems to be some measure of diffusion or scattering across the EASEL district for these people. In contrast, larger pockets of households listed in the '*Lower Managerial*' group appear to be shifted around the district with little change in the actual level of concentration. '*Intermediate Jobs*' are less concentrated across the EASEL district and a similar trend is realised for the '*Small Employers*' group, though clustering across contiguous OAs begin to appear. There are no significant trends realised for the '*Lower Supervisory*' group while for the '*Semi-Routine Jobs*' group, the pockets of concentration appear to be shifted around; this is also true for '*Routine Jobs*'. There are no significant changes for the '*Never Worked, Long Term Ill*' and '*Not Classified*' groups by 2021.

These are the major type of trends that will be observed when the scenarios are examined. All graphs and tables of statistics will be supported by catalogues of maps displaying the change in the spatial distribution of households over the 20 year simulation period. All map work is presented for the 2001, 2011 and 2021 years only. Overall, the baseline situation illustrates an EASEL district in need of intervention; there appears to be a trend toward segregation in some areas. When the population distribution is observed spatially, this trend is still evident; like households appear in clusters, though it is apparent that these clusters move from one district to another over time. The trends also highlight the immobility of households in low income socioeconomic classes as opposed to other households which appear to be more mobile. This suggests that the situation of deprivation and low diversity is likely to persist among low income households over time, justifying the claim that intervention is needed if improvements in the EASEL district are to be realised.

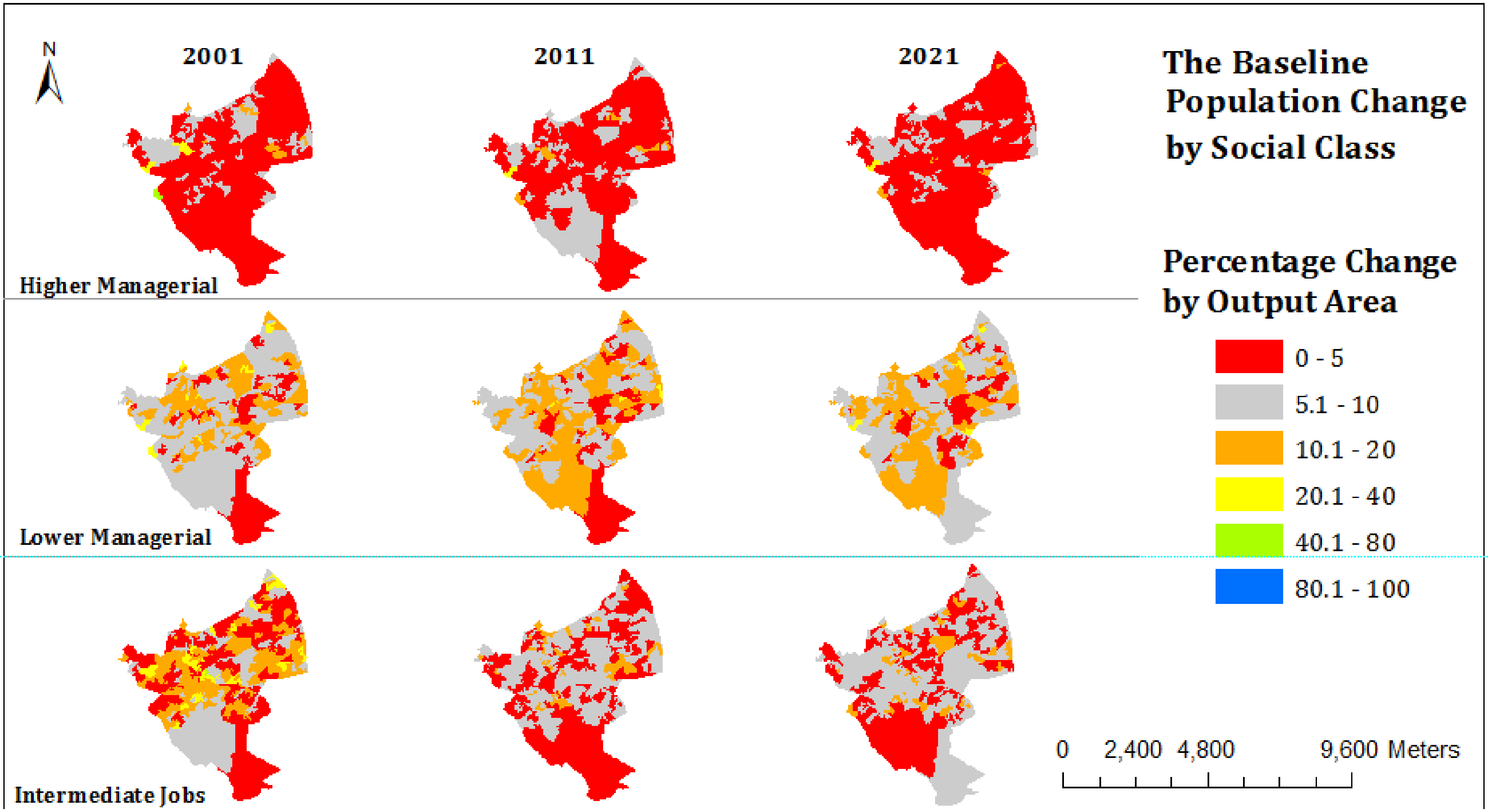


Figure 6.7a Comparison of Diversity Indices by Social Class over time (Baseline)

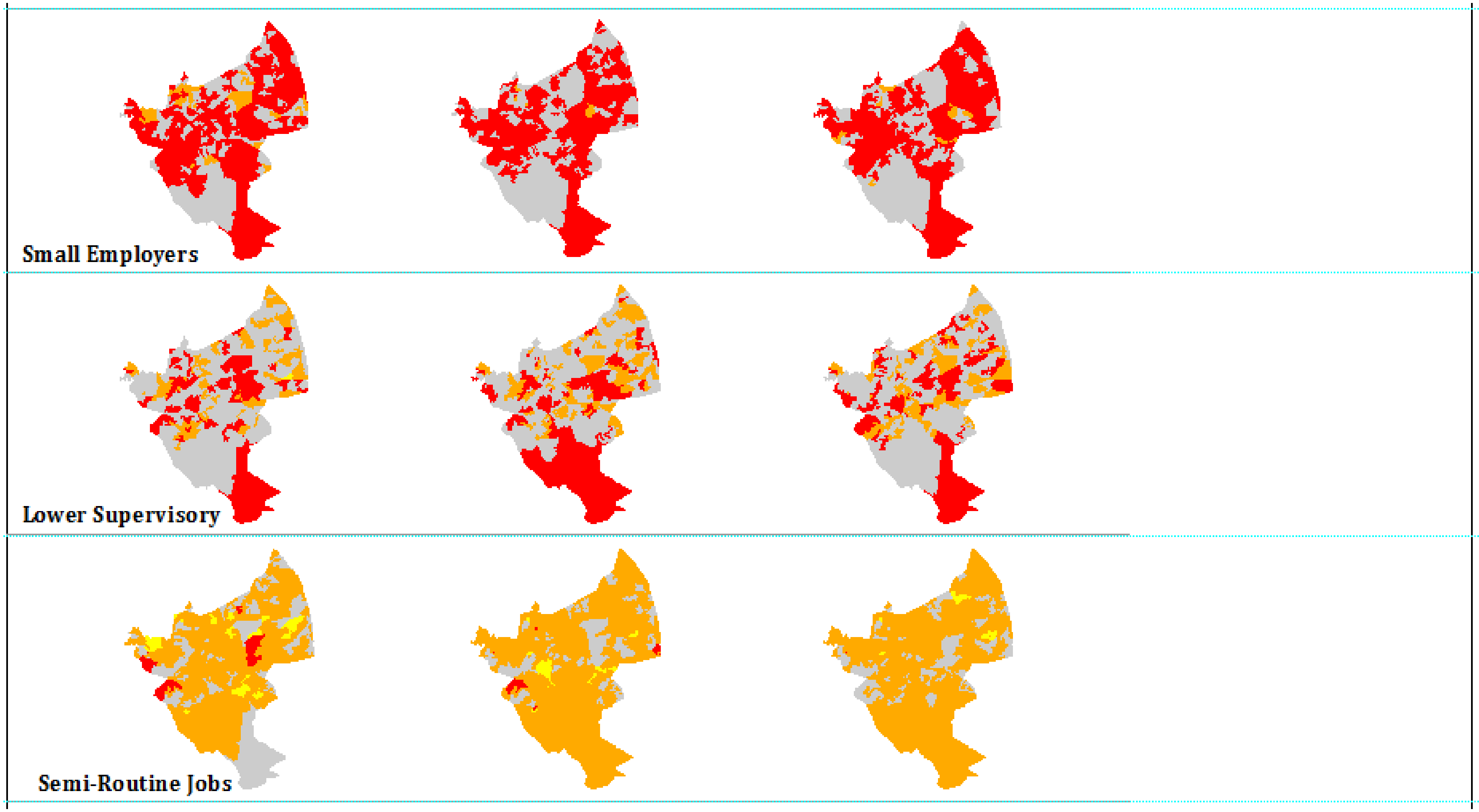


Figure 6.7b Comparison of Diversity Indices by Social Class over time (Baseline)

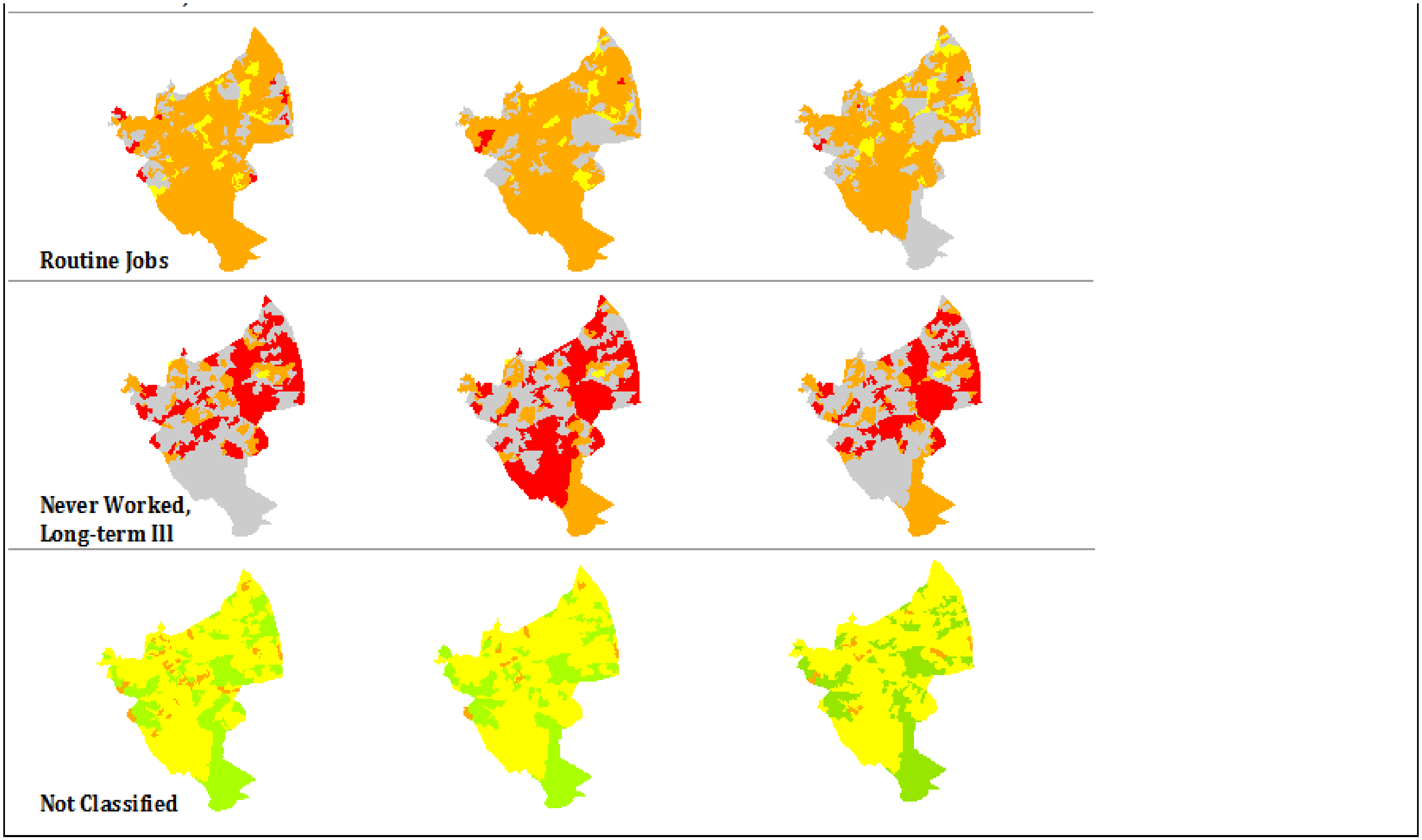


Figure 6.7c Comparison of Diversity Indices by Social Class over time (Baseline)

6.3 Describing the Scenarios and Results

The results of each scenario are compared and contrasted to the baseline situation over the 20 year simulation period. Plainly speaking the hope of the council is that the population distribution will change. However, in the worse case scenarios, the available housing will be taken up by the local community. Given the difficulty in predicting the national, or even city level response of the population changes in the housing market, these worse case scenarios are a pragmatic first stage in modelling the regeneration, and show the baseline from which the council needs to strive by increasing opportunities.

6.3.1 Scenario 1 Introducing a New Mixed Tenure Housing Development

As part of its regeneration programme, Leeds City Council has proposed to build a new mixed tenure housing development in the Gipton area of the Harehill ward, OA 00DAFM0025. Though Gipton is known to be an area affected by high incidences of crime, antisocial behaviour and other social problems, the DI for the chosen OA is noted to be approximately 79% in 2001; it is an area which is already mixed by socioeconomic class.

The nature of a mixed tenure development is such that households living in the development would be oblivious of the tenure make up of all neighbouring houses. In order to maintain the privacy of future potential householders the precise distribution of tenures associated with this area will not be discussed in this paper; a request made by personnel at the Leeds City Council (GJESSING, M., 2010). As a consequence, the details given in this exercise are approximations of the actual details. In total, there are to be 140 new homes built in the development area. The houses will be of varying accommodation and tenure types; terrace houses, semi-detached and detached homes as well as purpose-built flats. These new homes are to be sold to those willing to purchase homes on the private market as well as rented to social housing tenants in a 60-40% ratio respectively. Note that this ratio is outlined in the regeneration plans of the council. All houses in the development are to be similar in physical design though the size of the houses will vary. This is to ensure that there is no clear rubric that can be used to distinguish between housing tenures.

For the purpose of this model, however, 140 houses are added to the original house shapefile. These houses are representative of this new mixed tenure development. Like the actual development, the 60-40% model is used to govern housing tenure. The general layout used in the CHAIRS simulation is shown in **Figure 6.7** below. Note that this new mixed tenure housing development is a complete addition to the EASEL housing stock; that is, the number of other vacant houses is not reduced.

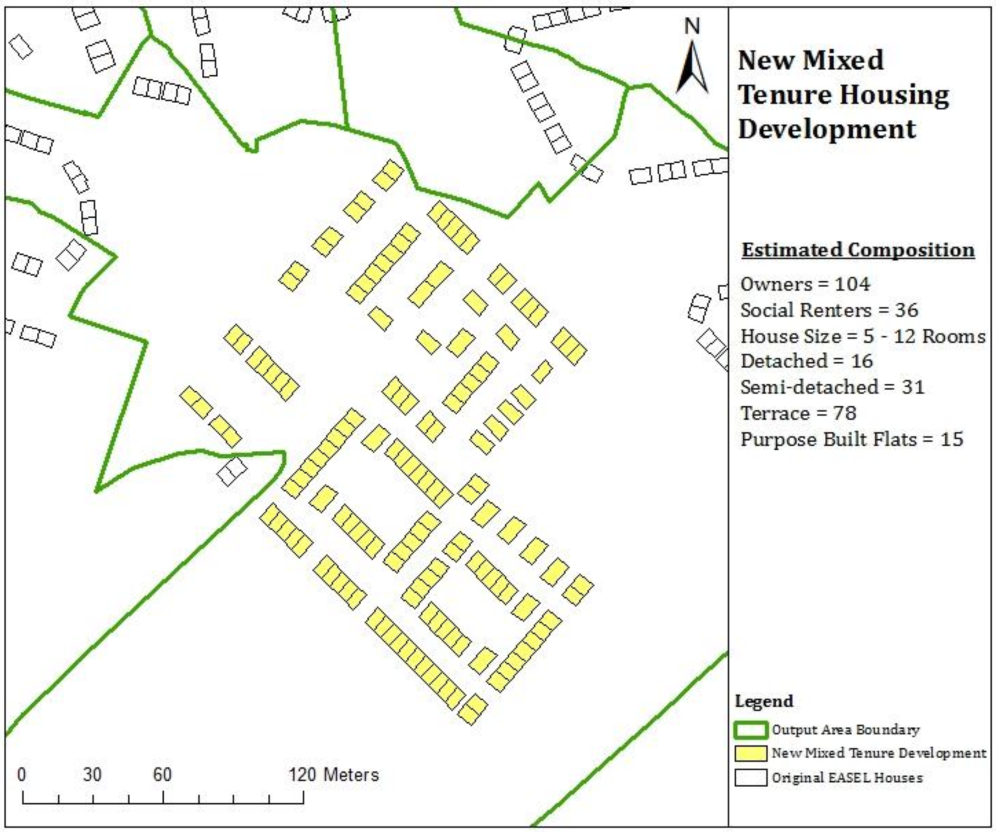


Figure 6.7 New Mixed Tenure Housing Development in the Gipton area

The results of the CHAIRS simulation will be analysed when this new mixed tenure housing development is added to the simulation. All results presented will be compared to the baseline situation to show the change in housing choice over time. A combination of total percentage change of population counts and diversity indices will be used to illustrate the results. The entire EASEL area will be examined for change, as well as the Gipton area specifically.

Out of a total of 286 OAs, there are 127 areas where the DI for 2021 is lower than the DI of 2001 when this scenario is applied. Conversely, there are 159 OAs that become more diverse over the 20 year period. Though this is difficult to illustrate logically in graphical form, such a trend points to the possible effects of residential relocation decisions; as some OAs become more and more diverse, others may become more segregated. The results of the Kruskal-Wallis test indicate that the baseline results and the results for scenario 1 at the 2011 period are independent of each other, suggesting a real change rather than just difference due to variation in the baseline run. In this way, the distribution of the population can be analysed with reasonable confidence.

The diversity indices for the Gipton area are of interest however. The graph below (**Figure 6.8**) compares the results of the baseline situation with the results from scenario 1 when the new mixed tenure housing development is built. With the addition of 140 houses, the results show a higher level of diversity in the Gipton OA over the 20 year period. Such a trend suggest that the new development area is able to attract a more diverse set of households as a results of the increase in mixed tenure housing. Without the policy, however, the baseline situation shows that the level of diversity in this area could possibly decrease over time, particularly between 2011 and 2012 where there is a sharp decline in diversity. Note that the starting conditions will be the same in all scenarios as each simulation begins with the same 2001 household distribution and it takes one year for the households to move into new houses.

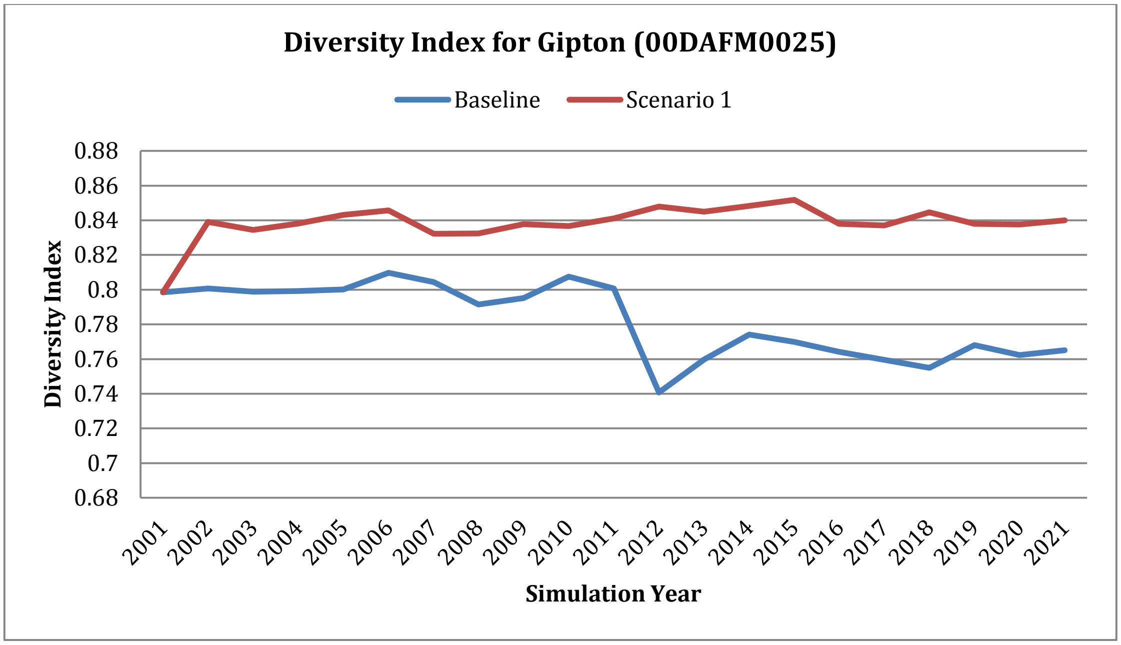


Figure 6.8 Diversity indices for Gipton area after scenario 1 implementation

Table 6.4 below is used to illustrate the types of households that have moved to the Gipton new development when the scenario is simulated. The table shows that moving from a total number of 0 in 2001, as a result of the new development, 19 of the *Higher Managerial* households occupying the Gipton area moved to the new development in 2011. This can be contrasted to the *Not Classified* group who in 2001 amounted to 46 households rising to a total of 68/67 households in 2011/2021 respectively. Of this group, ~21 households moved to the Gipton new development.

	Baseline	Gipton LLSOA 00DAFM0025		Gipton new development only	
Socioeconomic Class	2001	2011	2021	2011	2021
1 Higher Managerial	0	19	15	16	12
2 Lower Managerial	8	31	34	23	25
3 Intermediate Occupations	8	14	11	9	7
4 Small Employers	10	17	15	10	8
5 Lower Supervisory	9	24	38	16	26
6 Semi-routine Occupations	20	46	35	21	19
7 Routine Occupations	18	38	40	20	17
8 Never Worked, Long Term Ill	7	12	10	3	3
9 Not Classified	46	68	67	22	21

Table 6.4 Counts of households by socioeconomic status in Gipton new development area using scenario 1

Figure 6.9 and **Table 6.5** below are used to further explain the trends shown in the previous graph. Using the socioeconomic class variable, there appears to be a reduction of poorer households over the course of the simulation run when the 2001, 2011 and 2021 statistics are compared. For example, there are fewer households in the '*Never Worked, Long Term Ill*' and '*Not Classified*' categories as compared to most other categories where there are an increased number of households. Over time, the area loses almost 10% of those households thought to be in the more vulnerable socioeconomic categories (**Table 6.5**). This may suggest that this already diverse area does not favour poorer households. Such a claim may be further supported by the increased number of households in the higher socioeconomic groups; most notably the '*Higher Managerial*' category which jumped from 0% in 2001 to ~4.9% in 2021.

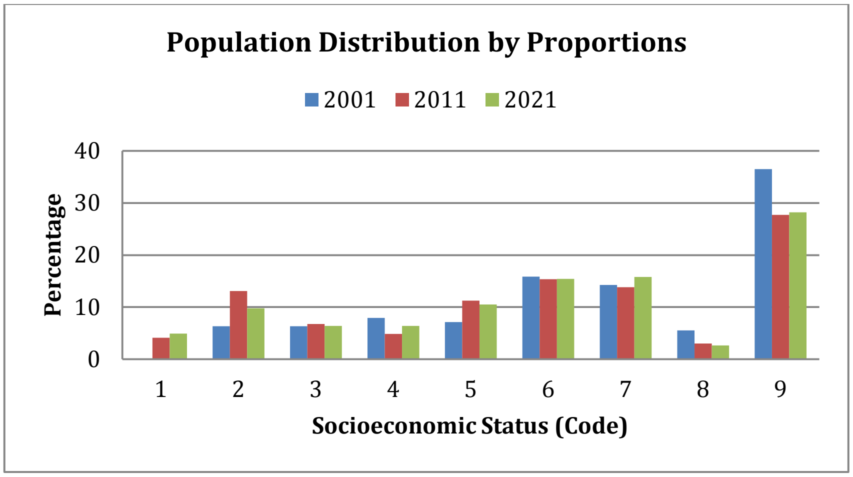


Figure 6.9 Population distribution of households by socioeconomic class by proportions in the Gipton new development area using scenario 1

	Percentages		
Socioeconomic Class	**2001**	**2011**	**2021**
1 Higher Managerial	0	4.1	4.9
2 Lower Managerial	6.3	13.1	9.8
3 Intermediate Occupations	6.3	6.7	6.4
4 Small Employers	7.9	4.9	6.4
5 Lower Supervisory	7.1	11.2	10.5
6 Semi-routine Occupations	15.9	15.4	15.4
7 Routine Occupations	14.3	13.9	15.8
8 Never Worked, Long Term Ill	5.6	3	2.6
9 Not Classified	36.5	27.8	28.2

Table 6.5 Distribution of households in Gipton by socioeconomic class using scenario 1 for the years 2001, 2011 and 2021

On the other hand, **Table 6.5** illustrates the possibility that a regeneration policy such as this can reduce or thin out deprivation in deprived areas; noting the reduction in households from socioeconomic class 8 and 9 over the 20 year simulation period compared to the increase in the socioeconomic class 1. This was discussed in **Section 2.4.3** which is based on the work of Uitermark (2003) as well as Atkinson and Kintrea (2000).

Similar trends can be observed when this data is presented spatially. The catalogue of maps below (**Figures 6.10a-c**) is an illustration of the change in the distribution of households over time when the socio economic status variable is observed. In general, the catalogue shows that over the 20 year period, there appears to be a reduction of the concentration of households across some socio economic groups while the converse is true for other groups. For example, for '*Higher Managerial*' and '*Lower Managerial*' jobs, there appears to be a convergence of these households in the vicinity of the new Gipton development. At the same time, the '*Higher Managerial*' group is more diffused over the entire EASEL district when the 2011/2021 maps are compared to the baseline situation at 2001. It is this diffusion across the EASEL district that is evident also in the '*Lower Managerial*' and '*Intermediate*' jobs categories; the evidence of change is particularly clear based on the change in the shading of the maps from dark to light where darker colours signify higher concentrations of households in each category. For the latter group there is a significant change in distribution of households in the bottom left of the map; over the 20 year period this large cluster of intermediate workers is reduced to levels below 5%. As lower levels of the socioeconomic ladder are examined more closely, there seems to be less change in the dispersal of these groups; for example, the '*Semi-routine Occupations*' and '*Routine Occupations*' categories. However by 2021, there appears to be a greater level of clustering among the '*Never Worked, Long Term Ill*' and '*Not Classified*' categories.

When these trends are compared to the baseline situation in 2011 and 2021, it is apparent that there are some differences in trends over some of the socioeconomic groups. When the higher socioeconomic groups are considered, there is a trend to move toward the Gipton new development area in both 2011 and 2021 as compared to the baseline situation in these years where this is not the case. These groups include the '*Higher Managerial*' and '*Lower Managerial*' groups. The converse is true for lower socioeconomic groups such as the '*Never Worked, Long Term Ill*' and '*Not Classified*' groups. Groups such as households with '*Intermediate Jobs*', '*Small Employers*' and '*Lower Supervisory*' appear to be shifted around the EASEL district with similar concentration levels.

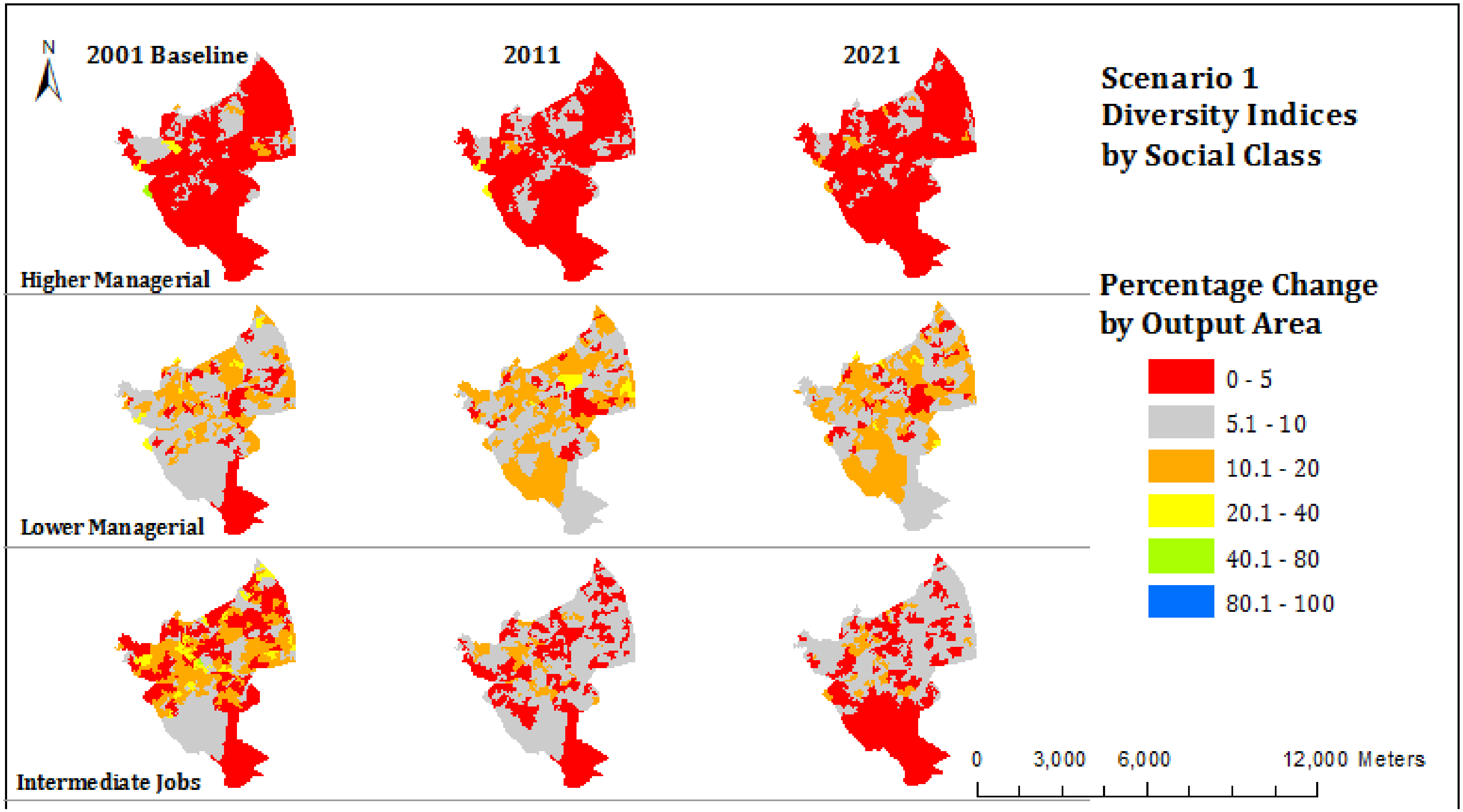


Figure 6.10a Comparison of Diversity Indices by Social Class over time (Scenario 1)

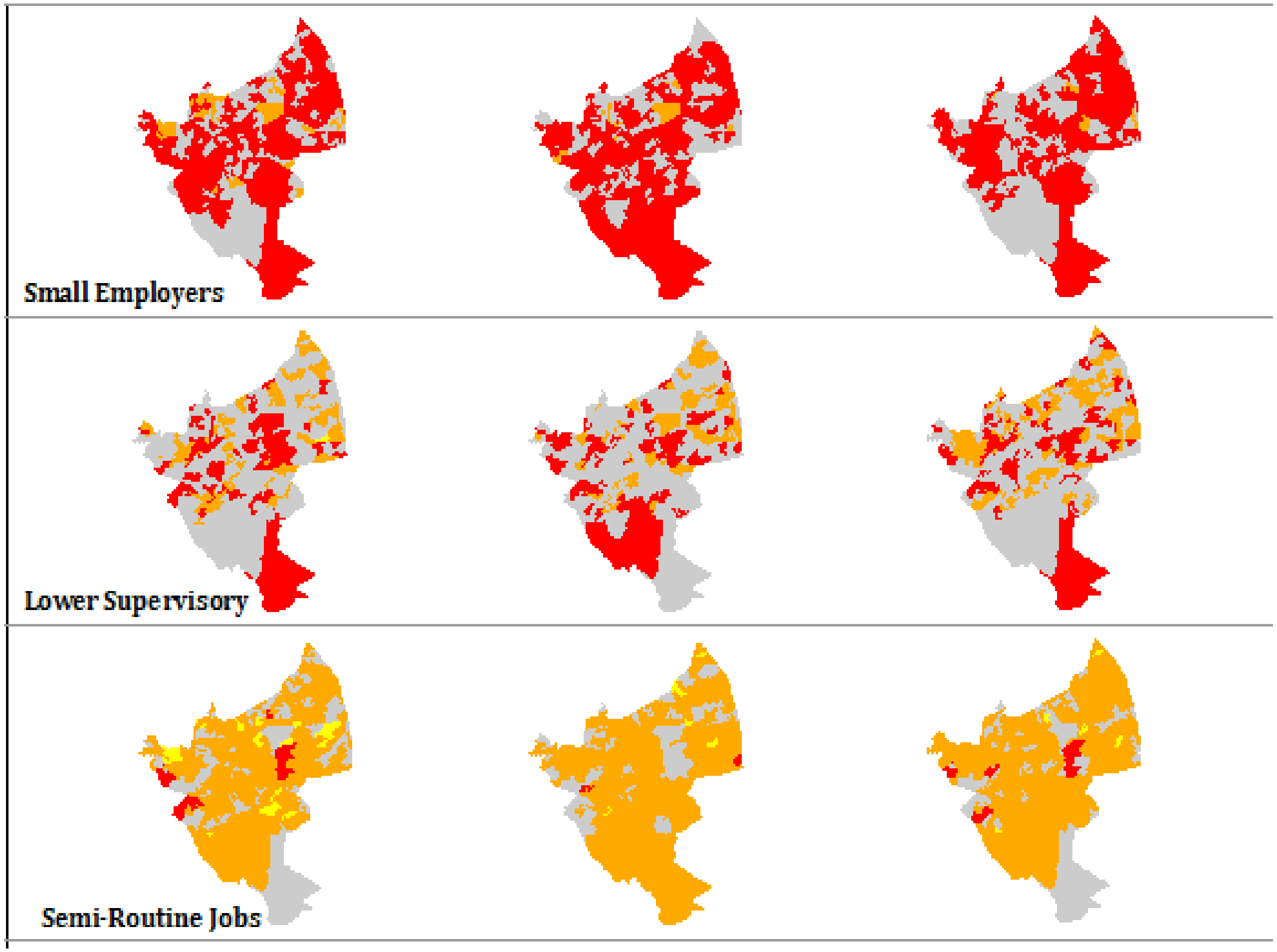


Figure 6.10b Comparison of Diversity Indices by Social Class over time (Scenario 1)

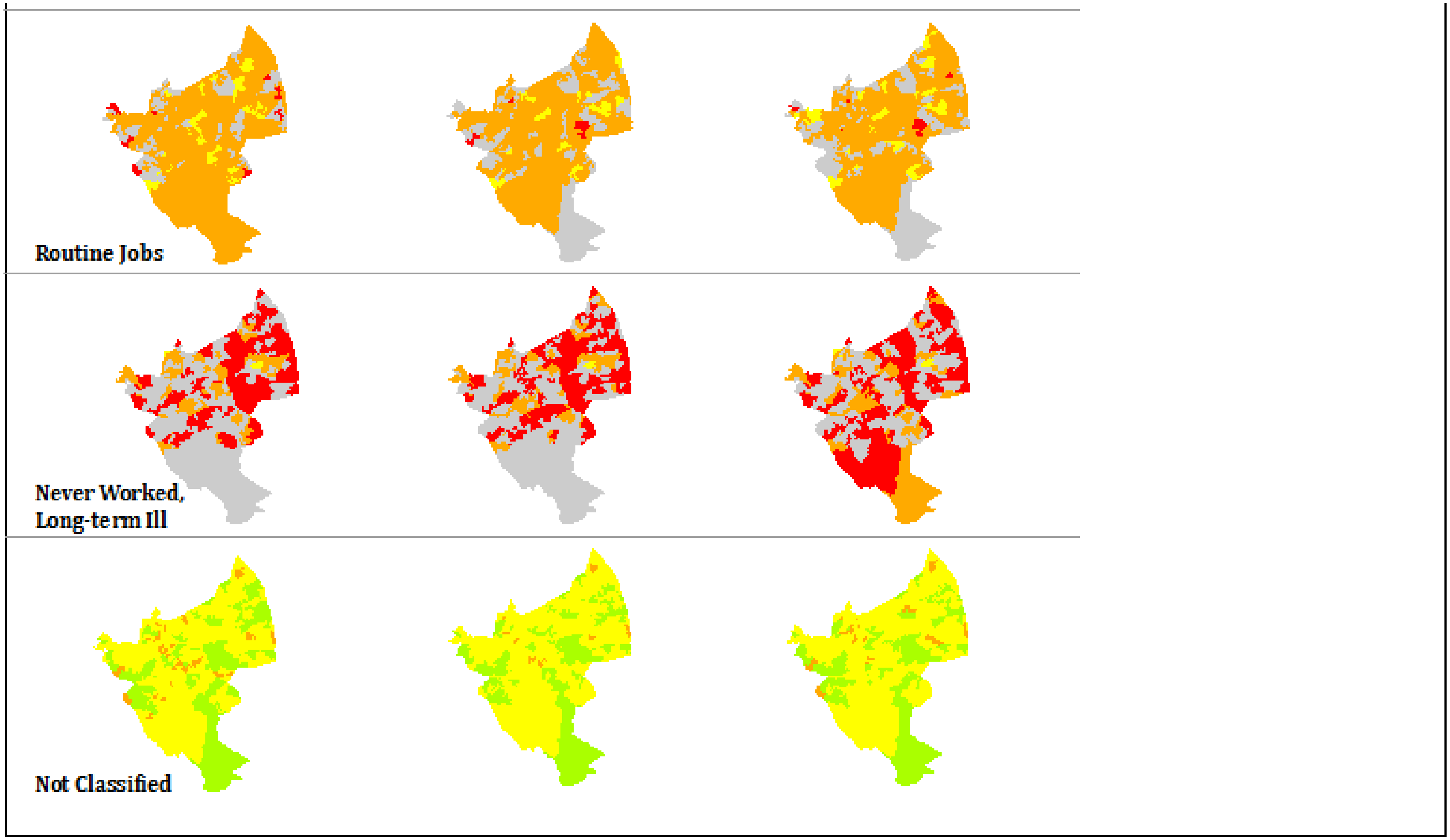


Figure 6.10c Comparison of Diversity Indices by Social Class over time (Scenario 1)

Based on the changes realised, scenario 1 appears to be successful in improving the socioeconomic mix in the Gipton new development area. This implies that by creating new mixed tenure developments, the diversity of a district can be improved, that is, the area is not just occupied by people already living in the local area. Such an improvement could make the area more attractive to investors and by so doing provide job opportunities for the unemployed. This in turn can lead to a more productive community, reducing deprivation and the incidence of crime and police costs (UITERMARK, J., 2003).

As presented in **Section 2.4.3** and based on the results of scenario 1, housing-led regeneration projects appear to be a viable mechanism for mixing communities by socioeconomic status as suggested by Tunstall (2003) and Bridge (2002). Furthermore, the results of the scenario suggest that there is a smaller concentration of low-income households indicating that deprivation could be reduced (KLEINMAN, M, 2000). Overall, this may lead to a reduction in welfare dependence as increased employment among householders may reduce the need for government assistance in the form of housing benefits.

However, the trend of gentrification also emerges from the scenario 1 results; as the Gipton new development area becomes more diverse, there is also a reduced incidence of the number of households in the more vulnerable socioeconomic groups – despite there not being a reduction in the amount of housing traditionally occupied by these groups. The general trend highlighted by the maps presented previously, suggest that segregation across these vulnerable groups appear to increase over the 20 year period. Though gentrification may be viewed as being positive for those living within the gentrified community, households of vulnerable socioeconomic groups are forced together because of their lack of economic power. This type of segregation can have adverse effects: proliferating poor quality housing, high rates of unemployment, high incidences of crime and antisocial behaviour all clustered in the same area (**Section 2.2.5**).

Whether the gains of diversity justify the negative impacts of gentrification is a question of much debate ((SEO, J. K., 2002); (DUTTON, P., 2003)). Gentrification can positively affect a rundown neighbourhood by causing significant improvements in the physical and economic structure of a community (CAMERON, S., 1992). As a result of this, house and rental prices may increase, making houses less affordable for low income households. Gentrification also changes the social structure of a community; gentrified areas tend to attract younger couples and professionals therefore eroding the family-type structure which previous low income households may have had (CAMERON, S., 1992). Not only are there demographic changes in this way, but racial minorities are also thought to be affected by this process. Side effects such as these are thought to erode community spirit and culture, intangible assets which evolve as a community grows without external intervention.

Despite these side effects, it is the process of gentrification which causes the improvements in the community through increased diversity. Thus the regeneration project appears to have beneficial effects in creating communities mixed by socioeconomic tenure. However, this is at the expense of low income households; those households which regeneration policies are thought to specifically target. This therefore suggests that housing-led regeneration policies may not be sufficient to assist low income households; rather supplemental regeneration policies may be needed.

6.3.2 Scenario 2 Changing the Road Network

For scenario 2, a new transport link directly joining the north and south of the EASEL district is created. Such a transport link is to allow for easier access to transport in general theoretically improving access to employment. The change in the road network will interact with the *Transport Rule* which directs households without cars to favour homes within one mile of a major transport link. In a practical sense, such a change in the road network would require road widening and other structural changes. Using the CHAIRS model, such a change merely requires a change in the aspatial attribute in the Roads shapefile from '*car walk*' to '*car walk majorRoad*'. This attribute determines road access.

The object of this scenario is to observe the changing diversity of the population highlighting specific changes which are effected as a result of this scenario. Based on the population behaviour highlighted in **Section 4.3.3.2,** when only the *Transport Rule* was executed, there was an increase in the number of households within a close distance of the new transport route. For this reason, the changing diversity indices of the 20 OAs nearest to the new road network are observed. However, overall, out of the 286 OAs in the EASEL district, reduced diversity was observed in 134 of these areas.

The physical layout of these areas is illustrated in the map below (**Figure 6.11**); the OAs surrounding the new road network are illustrated on the map using light green polygons. Here is where the change in the population distribution is more likely to change due to the close proximity of these houses to the new road network.

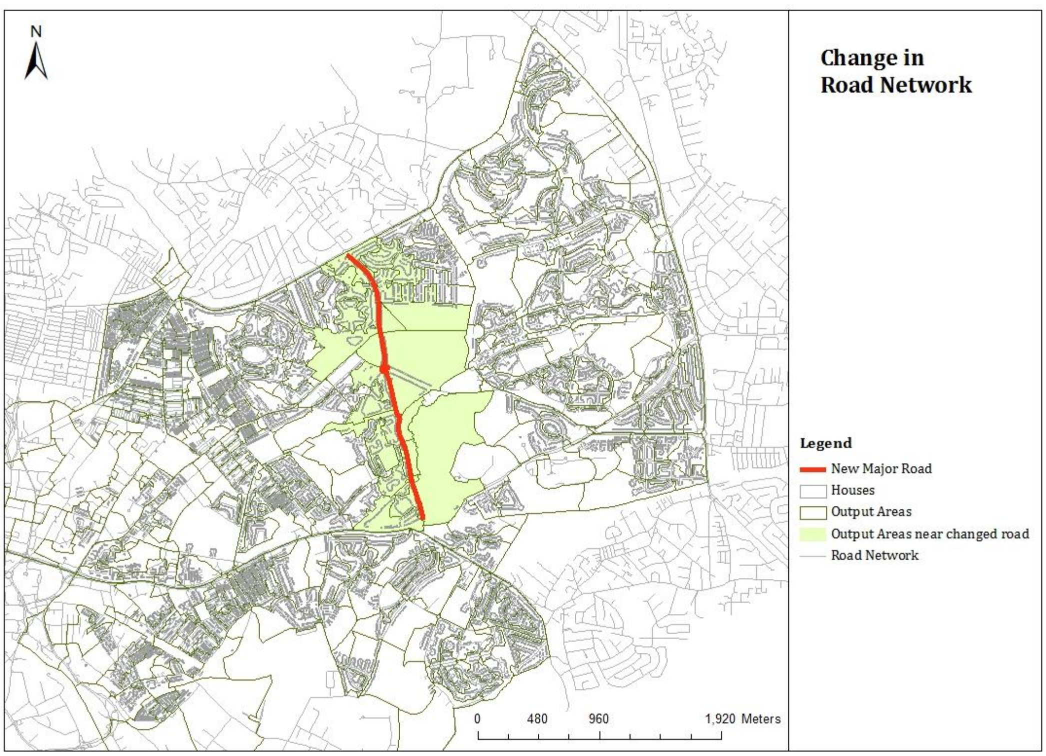


Figure 6.11 Map illustrating the change in road network and the surrounding districts

Diversity indices are used to compare the baseline situation with the 2011 and 2021 results. Noting that the baseline situation at 2001 is the same as the 2001 results for all scenarios, **Figure 6.12** below compares the diversity indices at the baseline situation in 2011/2021 with the diversity indices for the same time period under the scenario. Here there is a slight increase in the DI for the majority of the OAs noted. The areas which do not increase in diversity are all located in the Wykebeck Valley Road/Branders/Gipton Approach neighbourhoods of Gipton (**Figure 6.13**). The trend continues in 2021 when scenario 2 is compared to the baseline situation at 2021 as illustrated in **Figure 6.14**. Here there is a marginal increase in the DI for some OAs at this snapshot of time. If this scenario is to be used as a mechanism to increase the diversity of surrounding districts then the results suggests that scenario 2 does not cause any significant change in the diversity of the population between 2001 and 2021.

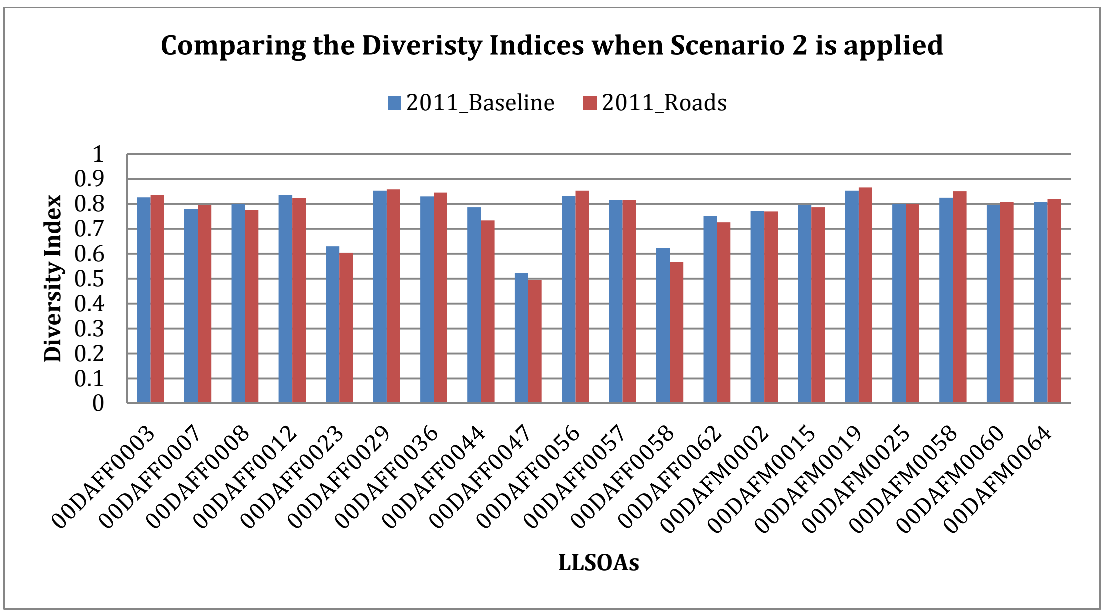


Figure 6.12 Comparing the 2011 baseline with the 2011 Scenario 2 results

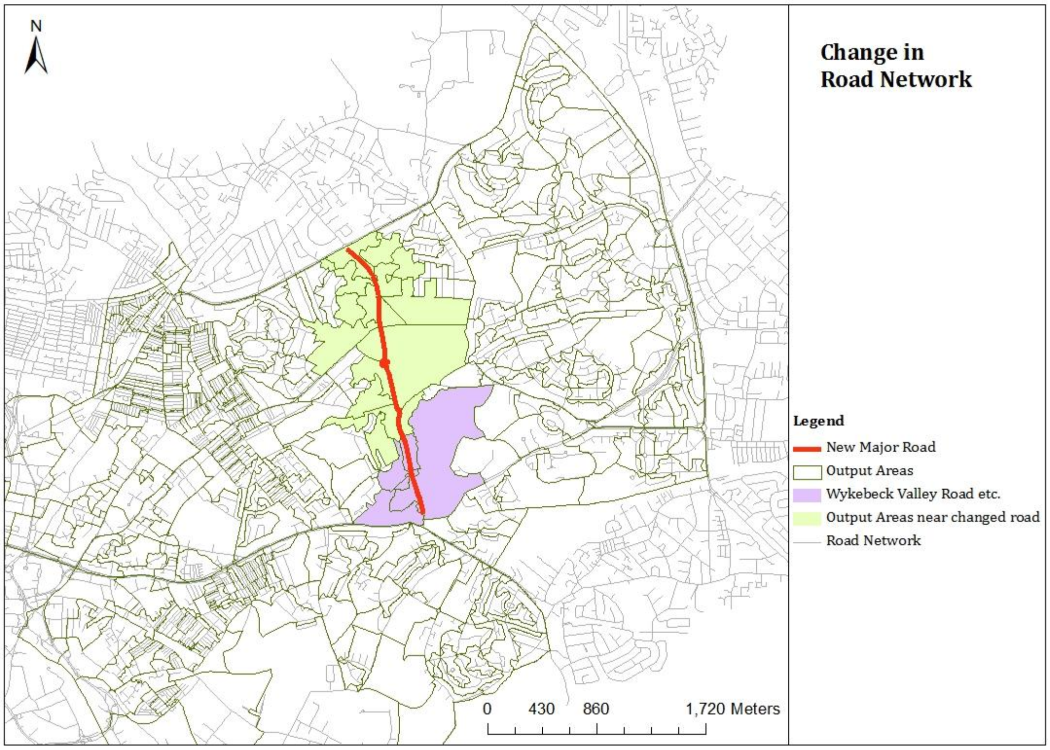


Figure 6.13 Wykebeck Valley Road/Branders/Gipton Approach areas highlighted in the context of the results for scenario 2. These areas do not increase in diversity.

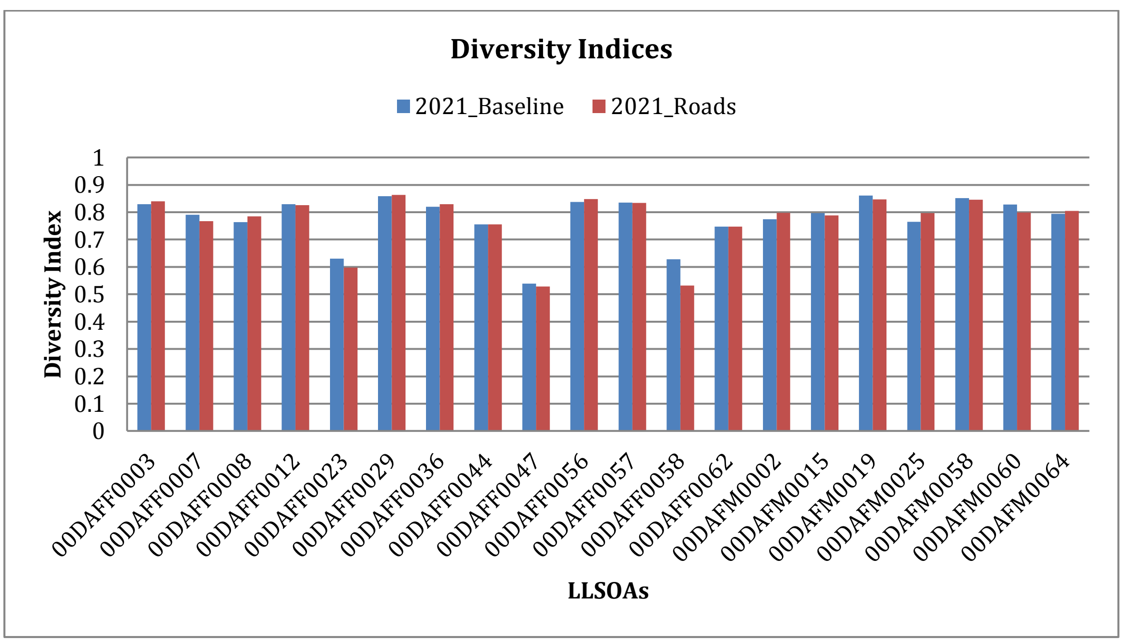


Figure 6.14 Comparing the 2021 baseline with the 2021 Scenario 2 results using diversity indices

Results for the Mann-Whitney U statistical test suggest that the null hypothesis cannot be rejected; that is, the two samples are not statistically different. This suggests that the diversity results for this scenario are not largely different when the 2011 baseline statistics are compared to the 2011 scenario results. A full table of results is be presented in **Appendix F, 6.1**. Note that these results were calculated manually for these two small datasets. Despite this observation, there are some considerable increases in the number of households in specific OAs of interest over the 20 year period. For example, in OAs 00DAFF0008 and 00DAFM0002, the population increased by 10% between 2001 and 2021. There are also marginal increases in 6 other areas. These changes are illustrated in **Figure 6.15** below and suggest that though the scenario does not increase the diversity of the areas surrounding the new road network, it has the potential to increase the size of the population in the surrounding neighbourhoods because the vacancies are taken up. Thus removal of vacancies in some areas may act to gentrify or exclude the poor.

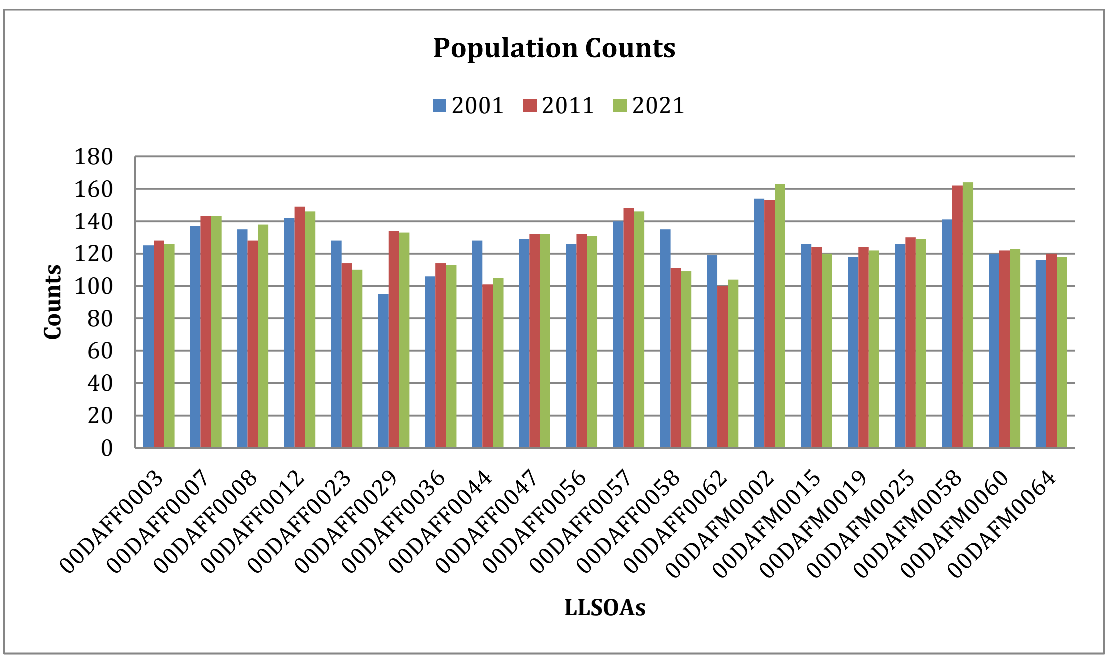


Figure 6.15 Population counts in the surrounding districts of the new road network after the implementation of scenario 2

The map catalogue below (**Figures 6.16a-c**) shows the change in the population distribution from 2001 to 2021. Notice for the '*Higher Managerial*' and '*Lower Managerial*' groups there is a trend for these households to cluster away from the new road network. Noting that the *Transport Route* rule focuses on households without vehicle transport it is possible that households in these groups already have access to cars. The fact that there are no major changes to the clustering pattern around the new transport route suggests that these households are not likely to have a need for easier access to public transport.

When the '*Intermediate Jobs*' and '*Small Employers*' are observed, there are no significant trends within close proximity to the new transport routes. However, there appears to be a levelling out of these population groups across the entire EASEL district. These trends may be consistent with the previously mentioned groups; the fact that such households are more likely to have access to their own transport implies that there is no preference to live near to the new transport route.

There is some concentration of households realised when the '*Lower Supervisory*' and '*Small Employer*' groups are observed. Also, when households in '*Semi-Routine*' jobs are observed, there is some clustering around the new road network as a result of the new transport route. Notice how the maps are shaded with a darker hue from 2011 to 2021; this is how the increase in the population counts around the new transport route can be identified. It is possible that these households do not have access to their own transport and are more likely to be affected by this scenario. This is not the same for the '*Routine*' jobs groups where a similar trend is thought to be expected, though a large cluster of this group is realised in the bottom left of the 2011 and 2021 maps. In a similar way, there are no significant changes around the new transport route when the '*Never Worked, Long Term Ill*' and ´*Not Classified*' socioeconomic groups are observed. This inconsistency of trends may suggest that the population growth may not be the result of the newly implemented policy.

These trends can also be compared to the baseline situation at 2011 and 2021. In general, there are no significant trends realised as a result of the changed road network though in some groups there is some diffusion as larger clusters recorded in the baseline 2011/2021 become smaller when the scenario 2 results are observed. This is true for '*Intermediate Occupations*', '*Semi-Routine*' and '*Routine Occupations*'. There is one recognisable change, however; for those in the '*Lower Supervisory*' group there is some measure of increased clustering around the new road network. Despite this observation, it is not sufficient to justify the effectiveness of this policy.

Overall, household behaviour is not thought to be significantly affected by the regeneration policy highlighted by scenario 2. Population growth in some areas is expected as a function of the modelling process, and this is balanced off by a reduction in the population in other areas, however, these changes are statistically indistinguishable from the baseline model variation.

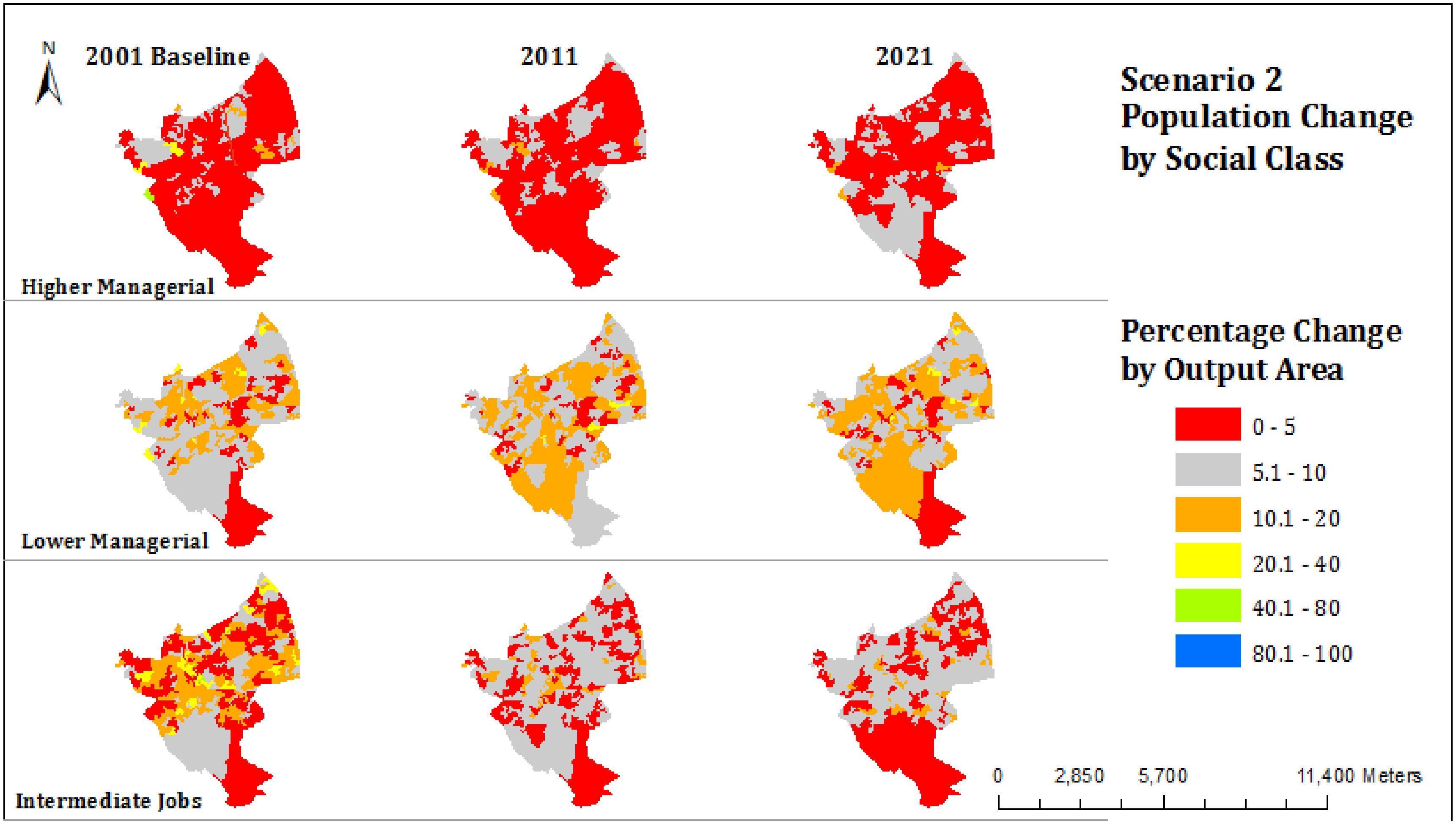


Figure 6.16a Comparison of Diversity Indices by Social Class over time (Scenario 2)

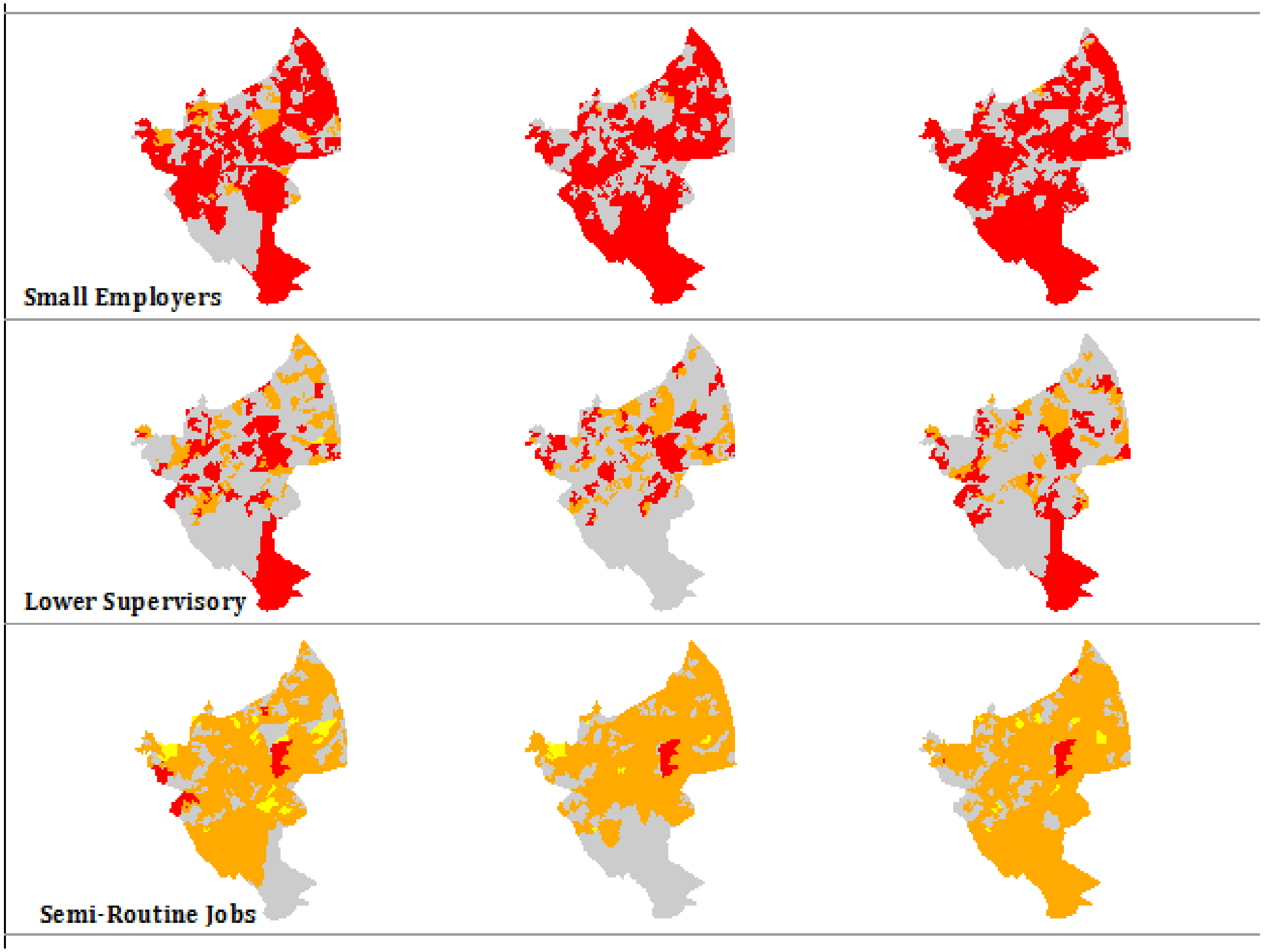


Figure 6.16b Comparison of Diversity Indices by Social Class over time (Scenario 2)

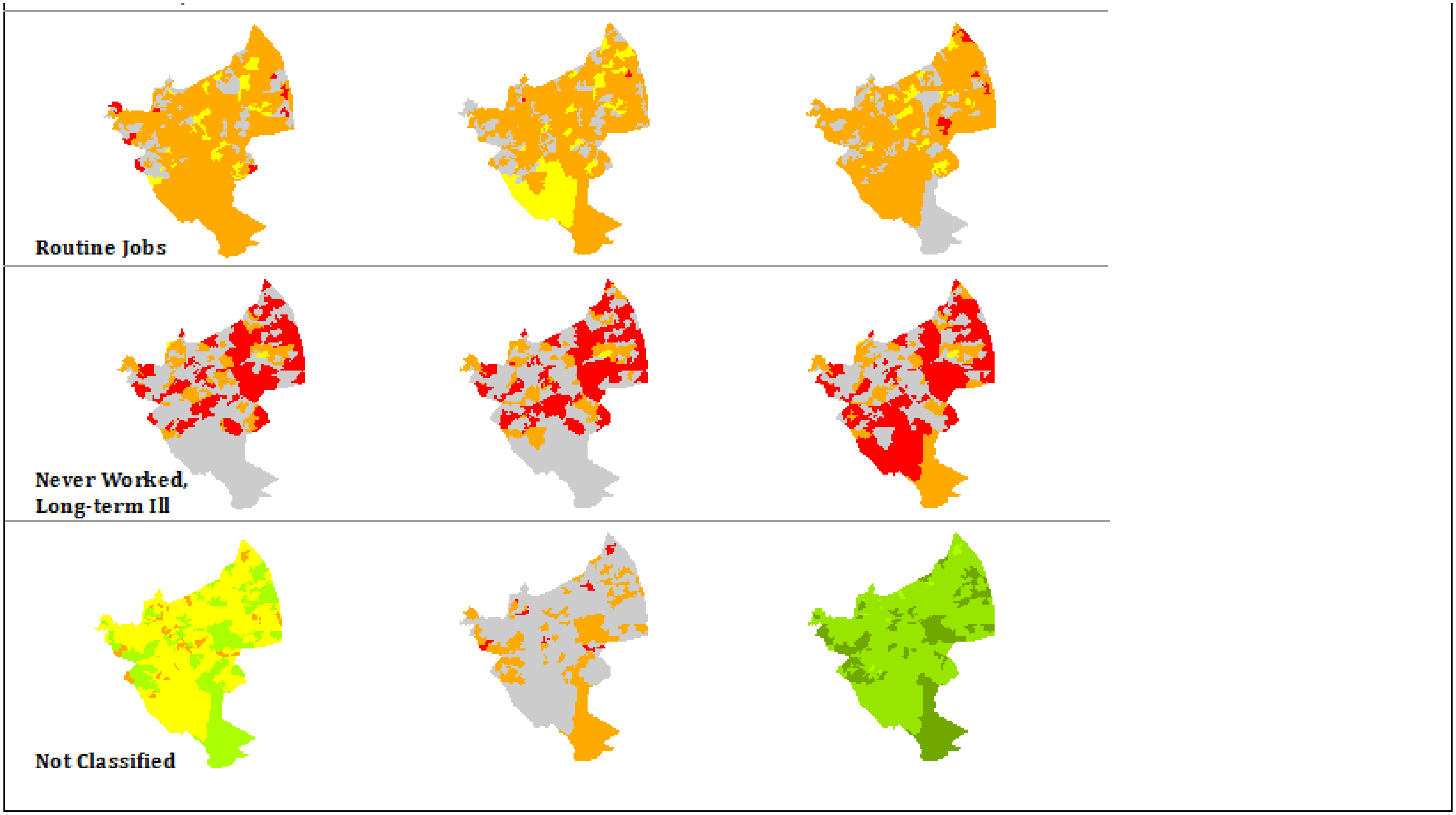


Figure 6.16c Comparison of Diversity Indices by Social Class over time (Scenario 2)

6.3.3 Scenario 3 Combining Scenarios 1 and 2

Scenario 3 is a combination of scenarios 1 and 2. Here the new mixed tenure development and the new road network are implemented simultaneously. The results are then observed. When the scenarios are implemented simultaneously there is a marked increase in the DI of the Gipton area where the new mixed tenure housing development was built. These trends are similar to those observed for scenario 1 and are illustrated in **Figure 6.17** below.

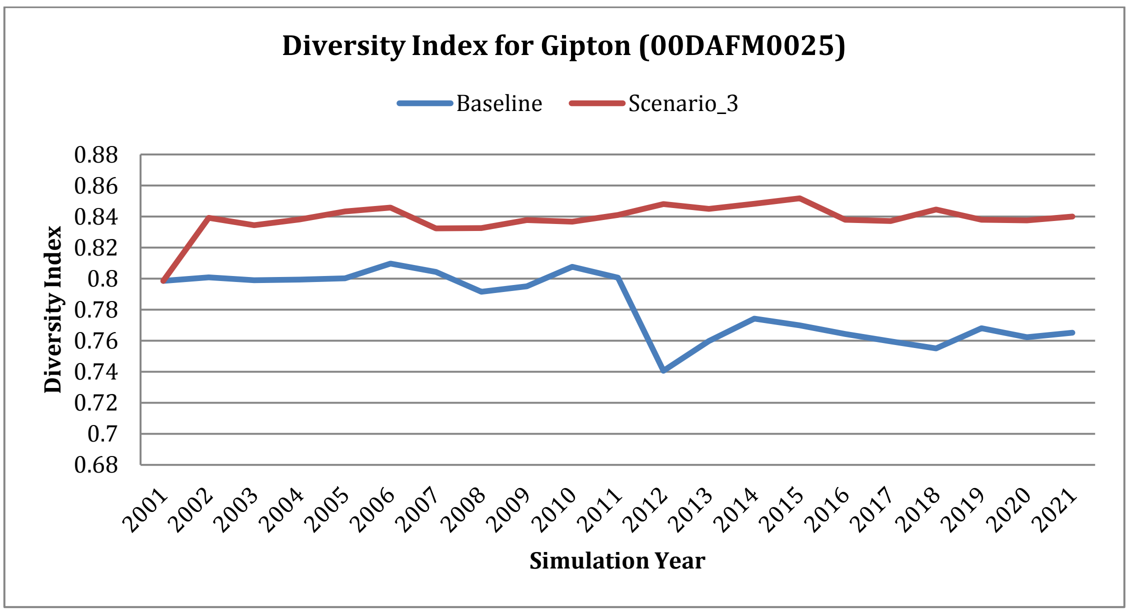


Figure 6.17 Diversity indices for Gipton new development using scenario 3

Like scenario 1, the trends can be analysed in more detail when the socioeconomic class is observed. **Table 6.6** shows an increase in the less vulnerable socioeconomic groups and a reduction in those socioeconomic groups thought to be more vulnerable. Much like the scenario 1 results, using the Kruskal-Wallis test, the results from scenario 3 are noted to be independent of the baseline results as at the 2011 period.

	Percentages		
Socioeconomic Class	**2001**	**2011**	**2021**
1 Higher Managerial	0	4.1	4.9
2 Lower Managerial	6.4	13.1	9.8
3 Intermediate Occupations	6.4	6.7	6.4
4 Small Employers	7.9	4.9	6.4
5 Lower Supervisory	7.1	11.2	10.5
6 Semi-routine Occupations	15.9	15.4	15.4
7 Routine Occupations	14.3	13.9	15.8
8 Never Worked, Long Term Ill	5.6	3	2.6
9 Not Classified	36.5	27.7	28.2

Table 6.6 Distribution of households by socioeconomic class in the Gipton new development using scenario 3

There is limited increase in diversity when the statistics are assessed with regards to the road network change, however, **Figures 6.18** and **6.19** help to illustrate this. This indicates that under current conditions, the scenario 1 element of this scenario is more effective in bringing about change in the EASEL area than the scenario 2 element where there is a change in the road network.

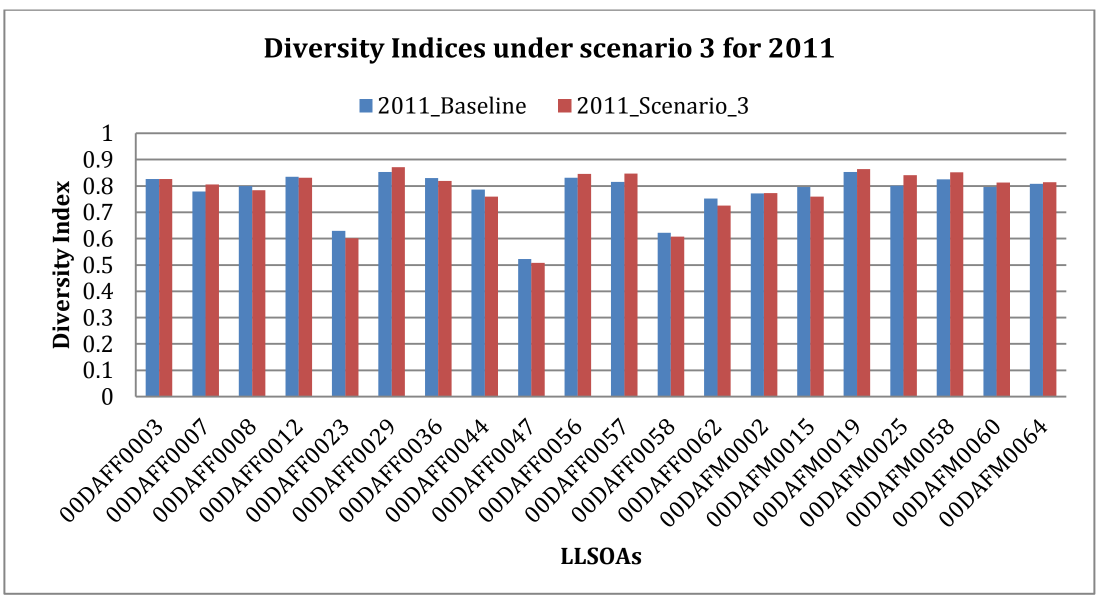


Figure 6.18 Comparing the 2011 baseline with the 2011 using scenario 3 results for the districts surrounding new road network

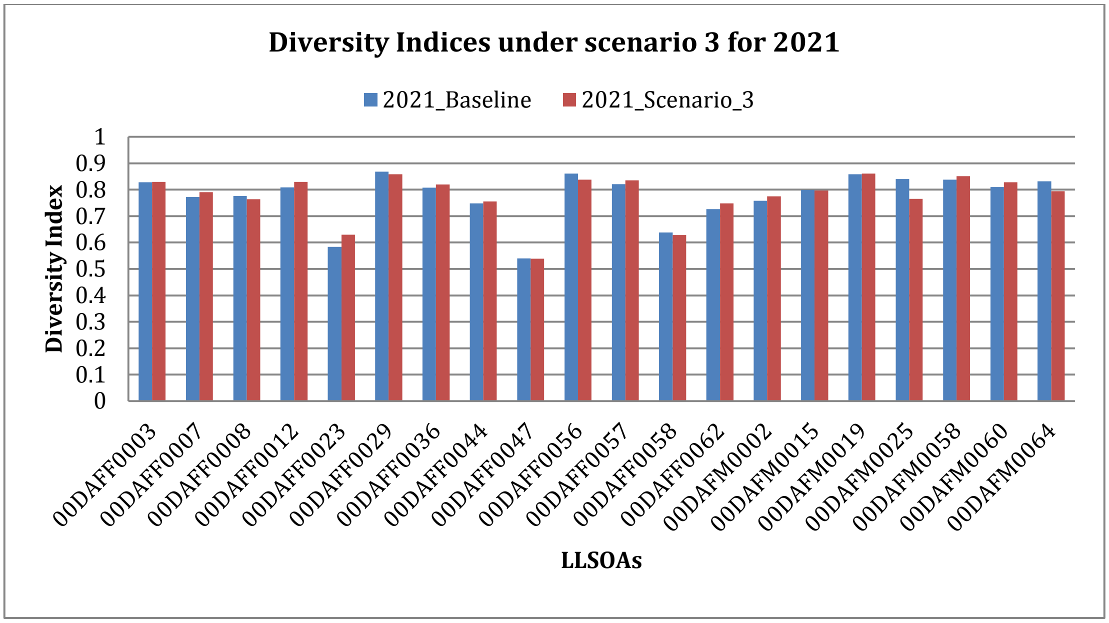


Figure 6.19 Comparing the 2021 baseline with the 2021 using scenario 3 for the districts surrounding the new road network

The catalogue of maps shown below (**Figure 6.20a-c**) show a trend for higher socioeconomic groups to gravitate toward the Gipton new development. These are the same types of trends observed in the results of scenario 1 and supported in the literature discussed in **Section 2.4.3**. Also, there are subtle changes related to the road network alteration, that is, there is a trend for those in low income socioeconomic groups to move closer to the new road network when the scenario 3 2011 and 2021 trends are compared to the baseline situation in 2011 and 2021. These groups include '*Semi-Routine*' and '*Routine Jobs*' as well as '*Never Worked, Long Term Ill*' and '*Not Classified*'. Though the results are subtle, this combination of scenarios appears to improve the immobility of low income households by providing easier access to transport links. These are households not likely to have access to personal transport.

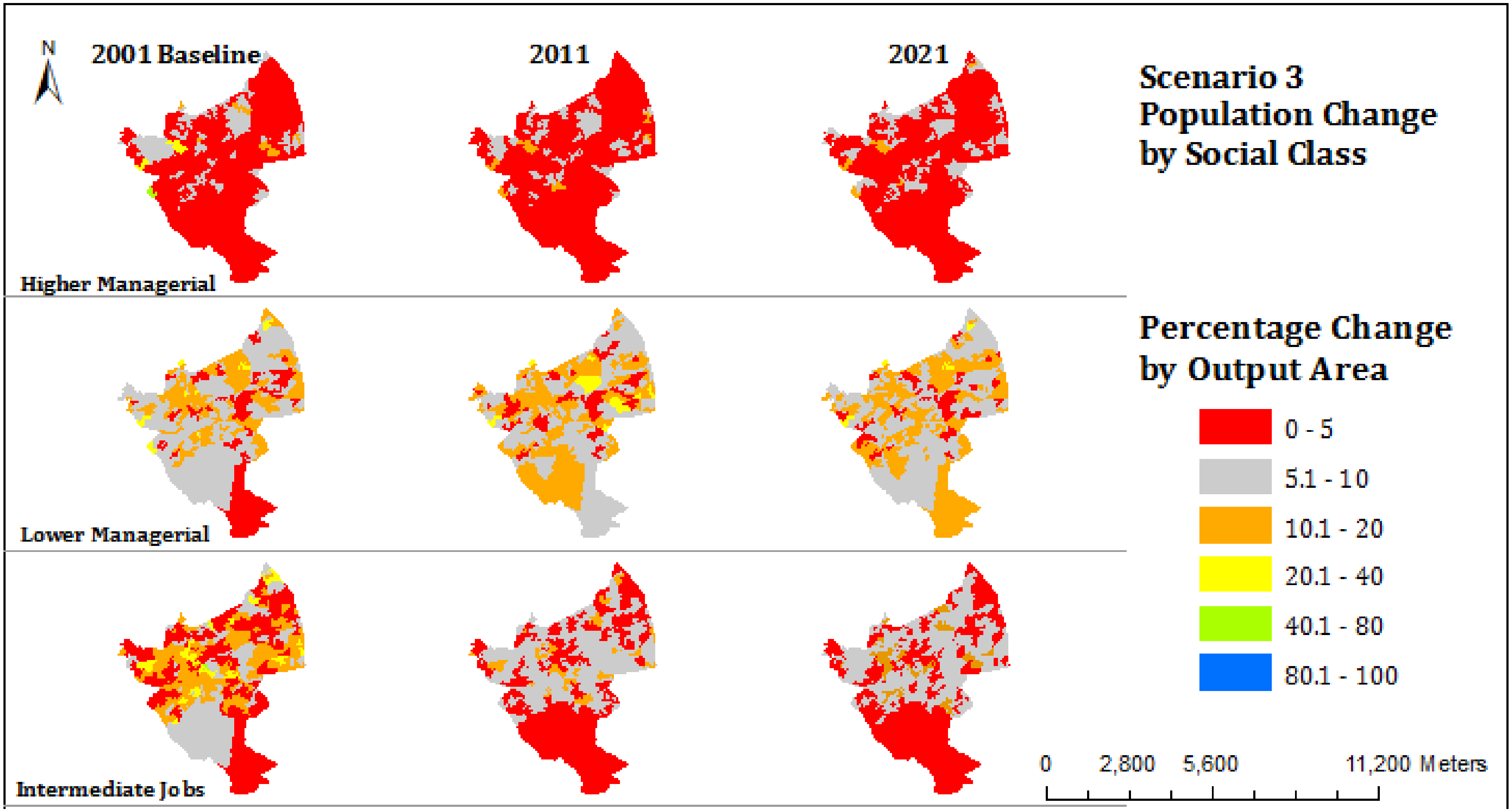


Figure 6.20a Comparison of Diversity Indices by Social Class over time (Scenario 3)

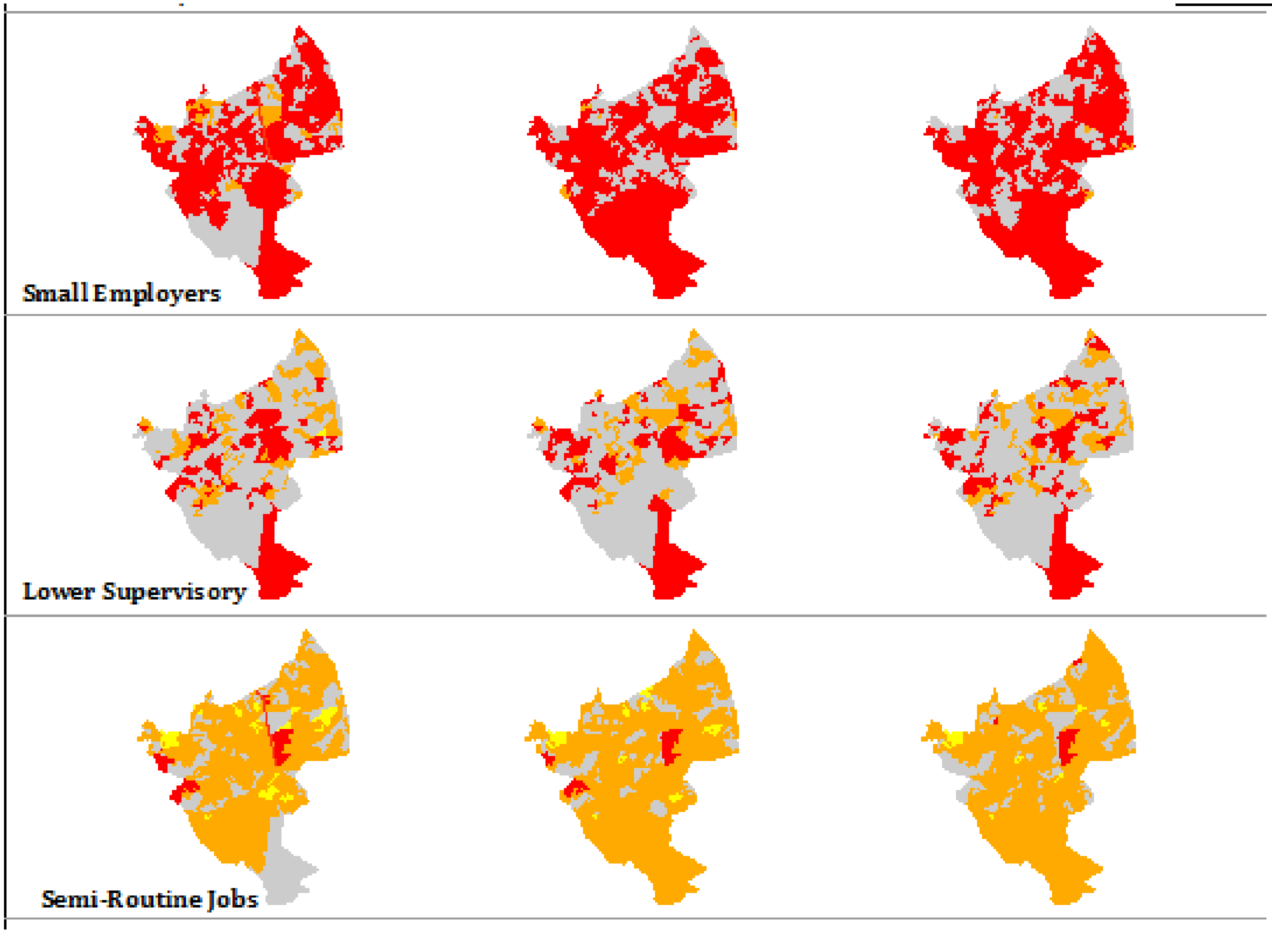


Figure 6.20b Comparison of Diversity Indices by Social Class over time (Scenario 3)

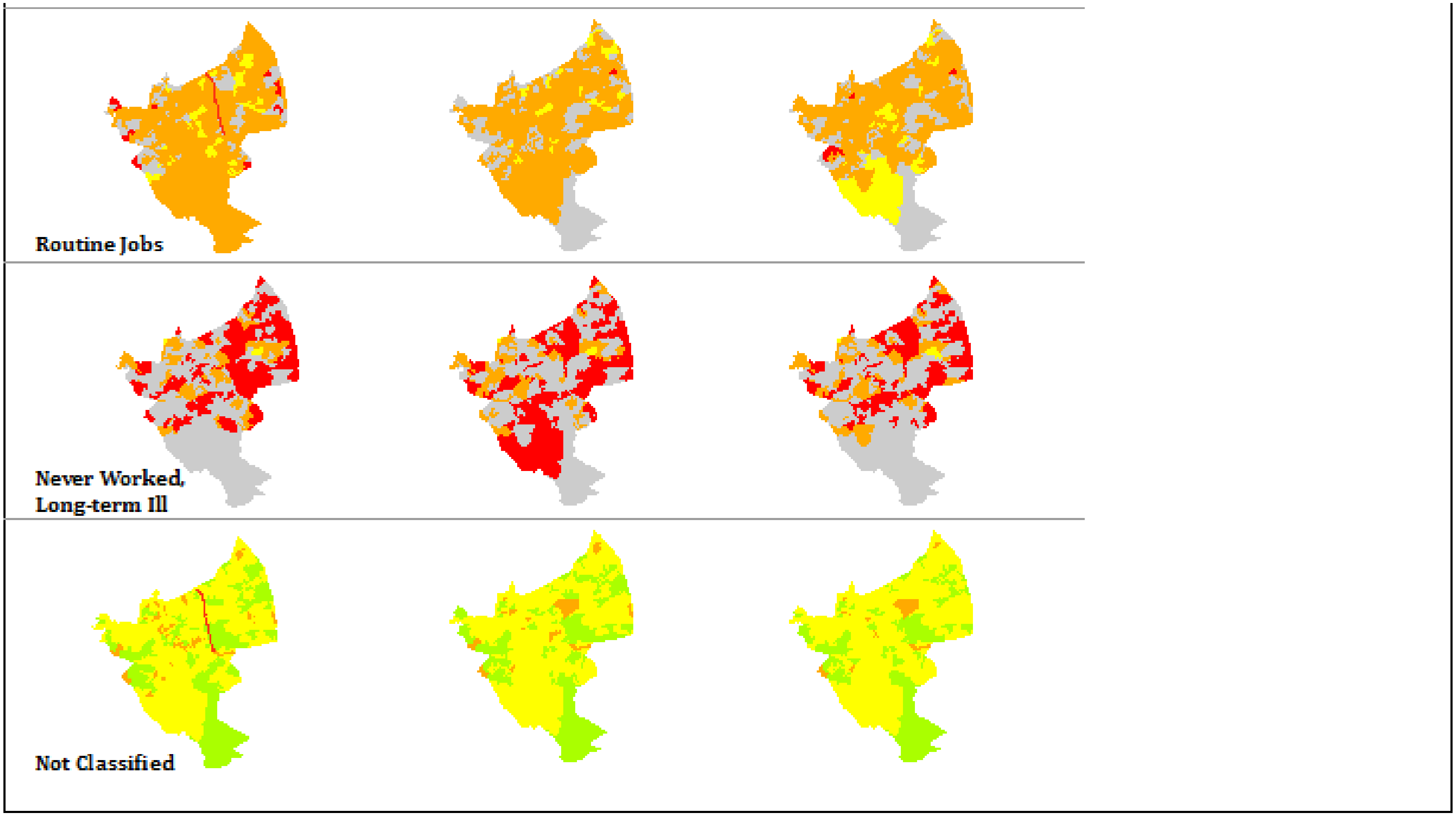


Figure 6.20c Comparison of Diversity Indices by Social Class over time (Scenario 3)

6.4 Further Discussion

Overall, the results of the CHAIRS simulation suggest that scenario 1 was the most effective of the three scenarios primarily because an increased level of diversity is realised when this scenario is implemented. Scenario 2 appears to make little difference when the diversity indices are analysed while only the effects of scenario 1 are realised when scenario 3 is implemented. Scenario 1 appears to increase the socioeconomic mix in the community, and this is the expected result hoped for by Leeds City Council. The potential effects of this are discussed in detail in **Section 1.4**. Thus, it appears that the presence of mixed tenure developments could indeed be a catalyst for increased diversity in the community in which the new development is built.

However, the results of the CHAIRS simulation (scenario 1 implementation) also suggest that gentrification may be a by-product when new mixed tenure developments are built. Here the results of this scenario are in agreement with Forrest and Kearns' (1999) view that regeneration projects involving tenure diversification have the potential to exacerbate social differences, potentially increasing social tensions as different social groups not sharing the same core values are brought together in one community. Thus it may be in the council's best interest to consider mitigating policies that could combat these negative effects should they occur. For example, the Leeds City Council may consider increasing the number of social housing options available in the new mixed tenure developments. **Figure 6.21** and **Table 6.7** below highlight the results of the CHAIRS simulation when the number of social housing options is increased. Here the model has been executed by implementing scenario 1 with 60% social housing options and 40% ownership options; the reverse of the original scenario. **Figure 6.21** suggests that over time there is an increased level of diversity.

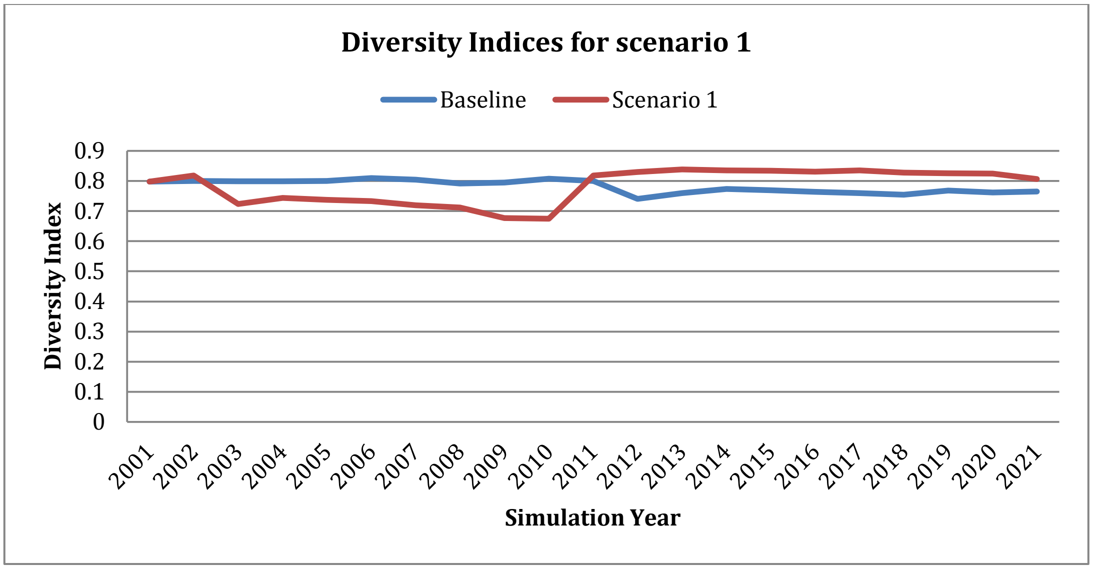


Figure 6.21 More Council houses allocated in new mixed tenure development

Unlike the original scenario however, where the number of households in the low income socioeconomic groups were reduced over time, **Table 6.7** shows that this is not the case when more social housing options are available. The table shows an increased number of households in higher socioeconomic groups but the reduction of households in low income socioeconomic groups is not as apparent. Thus it appears that with additional social housing options, the diversity of a neighbourhood could be improved without having to suffer the negative impacts of gentrification.

	Percentages		
Socioeconomic Class	**2001**	**2011**	**2021**
Higher Managerial	0	3.1	3.6
Lower Managerial	6.3	9.4	6.8
Intermediate Occupations	6.3	7.5	4.8
Small Employers	7.9	6.7	5.6
Lower Supervisory	7.1	7.5	8.8
Semi-routine Occupations	15.9	11.8	13.5
Routine Occupations	14.3	14.5	17.1
Never Worked, Long Term Ill	5.6	5.1	4.8
Not Classified	36.5	34.5	35.3

Table 6.7 Distribution of households by socioeconomic status in the Gipton new development area (Scenario 1 adapted; more social houses)

Amidst this adaptation, scenario 1 favours owners across the new mixed tenure development. Higher income households are the ones to occupy the new private homes and not low income households. Though the council hopes that by providing more opportunities for ownership, these mixed tenure developments would give low income households the opportunity to enter the private market, albeit that this is a model and not the real world, such an outcome is not evident in this model. By altering the ratio of social housing options available in the new mixed tenure development, the number of households in low income socioeconomic groups is less affected. Therefore, it may be in the council's best interest to consider such an alteration to the scenario 1 regeneration project in order to reduce the negative impacts of gentrification. Acting on these results would counteract the claims of researchers who claim that regeneration schemes such as these are purposefully used by the council to reduce government's stock of social housing (**Section 2.4.1**).

With regards to residualisation, the model does not show high levels of residualisation. Instead the population appears to be more evenly distributed when scenario 1 is implemented. Areas with low diversity indices in 2001 improve in diversity by 2021 as shown in **Table 6.8** below. Though highly integrated areas lose some of their diversity in this simulation period, the gains of integration in low diversity areas appears to explain this loss. Thus it cannot be said that the new mixed tenure housing development leads to residualisation of low income households instead, to some extent disadvantage is thinned out.

		Baseline		Scenario 1
LLSOA Name	**Output Area**	**2001**	**2021**	**2021**
Seacroft	00DAGE0048	0.37	0.40	0.48
Harehills	00DAGF0069	0.45	0.49	0.54
Gipton	00DAFF0047	0.45	0.54	0.54
Seacroft	00DAGE0011	0.49	0.6	0.5
Harehills	00DAGF0066	0.50	0.6	0.52
Richmond Hill	00DAGB0045	0.51	0.6	0.61
Richmond Hill	00DAGB0015	0.52	0.6	0.55
Gipton	00DAFF0023	0.54	0.63	0.58
Seacroft	00DAGE0012	0.6	0.61	0.6

Table 6.8 Change in diversity indices comparing the baseline and scenario 1 results

Also, recall that the IoS was earlier introduced. The resultant statistics for each scenario can also be observed over the simulation period. Note that the IoS (IoS) ranges from 0 to 1, where higher values represent more segregated neighbourhoods. The index is used to identify the level of segregation across the EASEL district between White households and Non-White households. **Figure 6.22** below is used to illustrate the change in the index over time for each scenario while the supporting statistics are presented in detail in **Appendix G**. In general, the model begins with a relatively low level of ethnic segregation as reported by the baseline illustration. Over time, the model appears to create more ethnic segregation during the 20 year simulation period. This is true for all scenarios.

The Seacroft district in the north east corridor persistently maintains high level of segregation throughout all scenarios. However, areas such as Harehills and Richmond Hill on the west side of the EASEL community continue to exhibit lower levels of ethnic segregation across all scenarios. Despite the tendency toward socioeconomic integration when scenario 1 is implemented, then, there are no clear trends toward ethnic integration as illustrated by the maps. The fact that neighbourhoods become more diverse when the socioeconomic variable is examined is a positive outcome. With an increased level of socioeconomic diversity, the welfare and social challenges arising within areas of concentrated deprivation may be reduced over time.

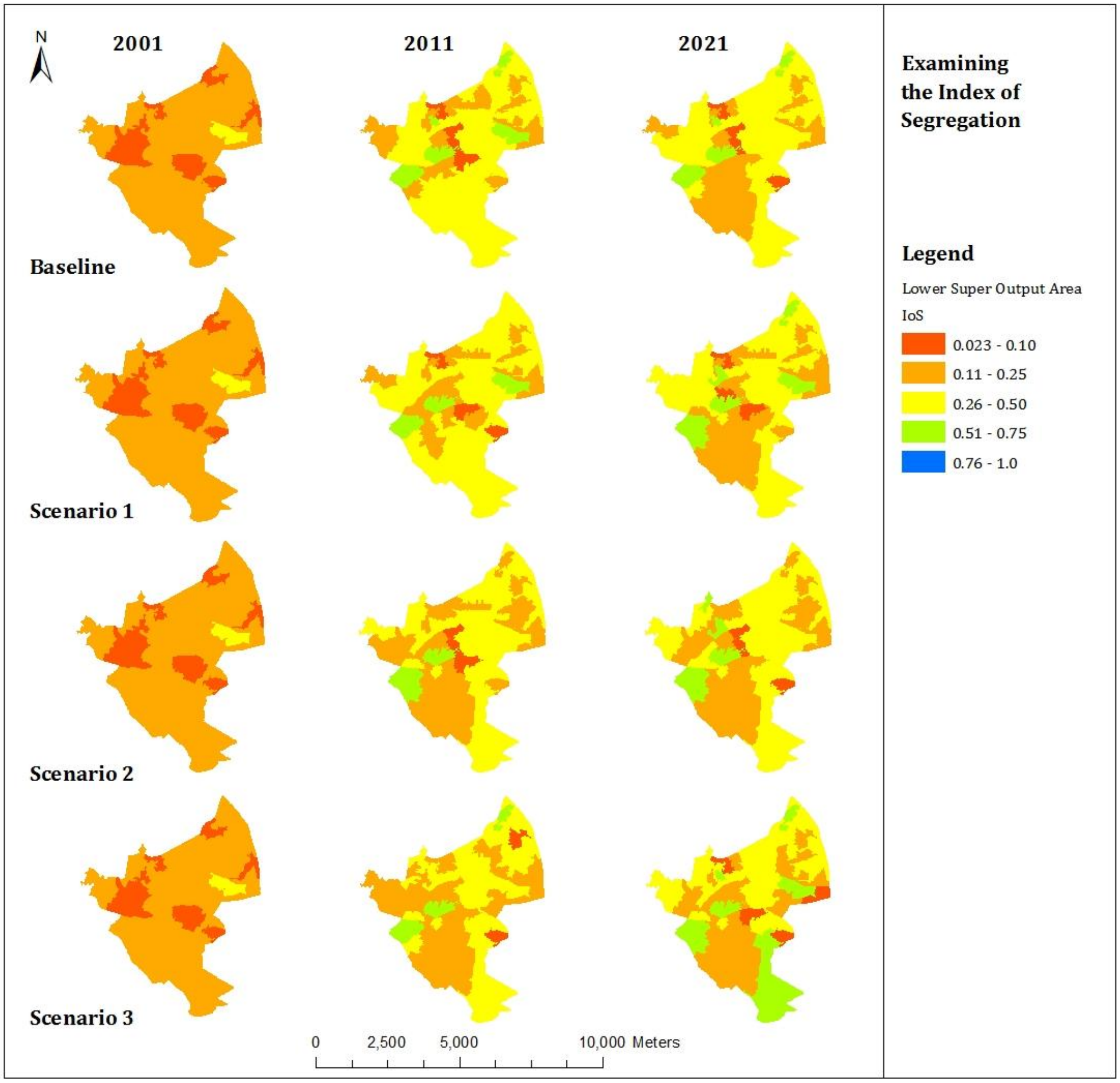


Figure 6.22 Examining the IoS 2001-2021 using the scenario 1 results

6.5 Conclusion

In this chapter the results of the CHAIRS simulation were presented. The results at the baseline situation highlighted that socioeconomic segregation is likely to increase if no policy interventions are made. Three policy scenarios were presented. Scenario 1 involved the building of a new mixed tenure development in the Gipton area. Scenario 2 allowed for a change to the road network to join the north and south of the EASEL district while scenario 3 was a combination of scenario 1 and 2. The results of the model suggest that scenario 1 was the most effective scenario in increasing the level of socioeconomic diversity, particularly in the area where the new mixed tenure development was built. The results of scenario 2 suggest that this regeneration scheme would have little effect on the diversity of the EASEL district if implemented. Though an increased level of diversity was created by scenario 3, it was apparent that this was the result of the effectiveness of scenario 1 element contained within this combination of scenarios.

Scenario 1 highlights the issue of gentrification. When this scenario was implemented a considerable number of low income households in the Gipton new development area disappeared from this community. However, when the level of social housing provisions were to be increased in this area, the exclusionary aspects of gentrification subsided. Though the new mixed development increased diversity, an observation such as this could highlight the need for the Leeds City Council to consider the policy further so as to reduce the negative impacts of gentrification. Finally, the model illustrated the likelihood of ethnic segregation to persist over time.

Chapter 7
Conclusions and Future Research

7.1 Summarising the CHAIRS Simulation Project

In the preceding chapters an approach to the simulation of residential mobility has been developed. **Chapter 1** was an introduction to the project. Here the topic of urban regeneration was introduced, the case study area of EASEL was presented and the challenges facing this area briefly discussed. Regeneration plans proposed by the Leeds City Council were also presented. **Chapter 2** discussed the subject of residential mobility and the behaviours that influence residential mobility. In this chapter, the family life cycle was noted to play an instrumental role in households choosing to relocate while financial constraints, family size, neighbourhood quality and access to goods and services were important considerations when a new house was to be chosen. The chapter also outlined the influence of housing policy in shaping the past and present communities. Housing policy is noted to be one of the mechanisms which have contributed to the problems of concentrated deprivation across some Local Authority estates, namely through residualisation. The chapter contrasts the arguments surrounding regeneration, noting that regeneration policy has the potential to create mixed communities though gentrification may be a consequence.

Chapter 3 examined the quantitative method by which the research question would be examined. The chapter discussed the value of computer modelling while presenting different types of modelling techniques that could be used to answer the research question. ABM was selected as the technique of choice to create a residential mobility model of the housing market using the behaviours identified in **Chapter 2**. The model, called CHAIRS, was created by implementing seven residential mobility behaviours. These included behaviours based on households' finances, access to transport routes, access to schools, ethnic tolerances, known areas and neighbourhood quality.

In **Chapter 4** the methodology used to build the model was presented. Inputs such as shapefiles used to represent houses, OAs, schools and roads were illustrated. Individual level household data created by the process of microsimulation was also presented. The model process was discussed while algorithms to represent the inner details of each behavioural rule were presented along with verification details which proved the correct working of each rule individually. Model outputs and assumptions followed this discussion.

Chapter 5 tested the realism of the model. Here the model was calibrated and validated using real world survey data based on the EASEL district from Acxiom's ROP. Each behavioural rule is used as a parameter in the model. Various combinations of the rules were executed in such a way that each set of rules could be compared to the results of the ROP. The calibration/validation process is made up of a series of reweighting and comparison exercises where the counts of households from the model results are compared to counts in the ROP. The combination of rules which generated the least number of errors was selected as best matching the actual real world situation reported by the ROP. This rule set combination comprised of 3 rules; *Transport*, *Ethnicity* and *Socioeconomic class*.

In **Chapter 6** the selected rule set resulting from the calibration/validation process was used to generate results when the scenarios were implemented. Three scenarios were implemented: a new mixed tenure development, a new transport link and a combination of these two scenarios. The results of the model showed that when a new mixed tenure development is built in an existing neighbourhood it has the potential to increase the level of diversity in the area. However, low income socioeconomic groups become excluded from the area, despite the fact that in total the number of houses increases and the baseline population is static, that is, there is no reduction or increase of any household types – as the population does not evolve – but rather a pervasive exclusion from the area. This is surprising given the increase in housing. Further analysis of the model was able to highlight the fact that if more social housing provisions are included in this new mixed tenure development, this effect could be reduced. If the original combination of tenures are to be used in the new mixed tenure housing development as planned by the Leeds City Council, the results of the model suggest that this new mixed tenure development is not sufficient to reduce the negative impacts of regeneration and therefore complementary policies may need to be considered. Scenario 2 was not successful in effecting a significant change in the distribution of the population while scenario 3 was only effective when the population distribution in the new development area was considered.

Overall, the model was used to represent residential mobility behaviour, implement regeneration policy and forecast the social mix over time. The results suggest that some regeneration policies *can* lead to socially mixed communities. Based on the behaviours represented in the CHAIRS simulation, new mixed tenure developments are more likely to improve socioeconomic diversity than changes in the road network. A by-product of this is the thinning out of deprivation as the model illustrated diversity indicators can improve across the district over the simulation period. *If* a socially mixed community is a sustainable community then it can be concluded that housing-led regeneration projects can lead to sustainable communities. These are the insights that can be gained when ABM is used to simulate residential mobility and regeneration policy.

7.2 Reflections and Further Research

The creation of a model such as the CHAIRS simulation does not come without challenges. These challenges in the design and implementation of the model ranged from correctly quantifying soft variables, extending the design of the model to increase behavioural dynamics and obtaining useful individual level data, validating model outcomes using the ROP dataset through to adequately utilising limiting computer capacity.

By nature qualitative literature lacks the precision which is indicative of quantitative analysis. Thus, the qualitative literature may report that households move short distances but in order to use this information within a model, the term 'short' must be translated into an actual distance. The challenge here is determining the correct definition for terms such as these so that they can be usefully incorporated in the model. Through optimisation procedures, more precise figures can be identified for use in the model. **Section 5.3.3** discussed the optimisation procedure, identifying more precise figures to be used for the distance parameters needed.

Several comments have been made on the closed nature of the system. That is, the demography of households do not change throughout the lifetime of the model. In a similar way, there is no in/out-migration while model details such as neighbourhood quality and housing unit layouts remain static over the lifetime of the model. A more individual, rather than household-based model might also clarify and improve decision-making. In a similar way, individuals, as opposed to households, could be modelled so that the family life cycle could be more clearly represented. Ideally, more dynamic demographic processes, such as these, would be included, however such models are sufficiently complicated as to form full research projects in their own right. In terms of further work, a demography model could be incorporated into the present model. In this way, births and deaths would be accounted for. This would result in households growing and shrinking, influencing decisions related to the number of rooms required in a house, access to schools and neighbourhood quality; all significant for households. By increasing these dynamics in the model, the goal of moving house for a more appropriate situation could be more effectively managed in the behavioural rules.

Residential mobility is a complex process. **Section 2.2** outlines the major drivers influencing this complex process. However, based on the results of the calibration/validation procedures presented in **Chapter 5**, only three of the seven behavioural rules implemented were used in the final simulation. One might expect that all behaviours developed should be included in the final model, though it seems equally likely that different households use different combinations of rules to find a new house. Considering this, further exploration could be carried out to refine the behavioural rules. For example, age specific behaviours could be examined, preference thresholds by ethnic group could be identified, buffer zones used for the *Known Areas* and *Schools* rules could be further examined and optimised. Likewise a more detailed social housing model could be implemented to replicate how social housing tenants move to and from the new intermediate market. At the moment, the model replicates a situation where social housing tenants are directed toward the private sector rental and private ownership markets progressively as governed by the *Tenure* behavioural rule. This is broadly in line with government policy, but ignores many of the nuances in the system, not least the processes by which those remaining in social housing move or are moved around the city. In addition to a more fully-formed social housing decision-making model, other mobility drivers such as estate agents and vacancy chains could be included in a future version of the model thus increasing the level of real world dynamics. Likewise, a mechanism to change environmental conditions such as the IMD for each OA could be included and house prices could be added along with the impact of changing interest rates. Again, these are not insignificant modelling areas in their own right.

It should be noted, however, that increasing the functionality of the model does not automatically mean that it would reflect reality more accurately; depending on the metric used for validation, a model of limited system components can often do a perfectly reasonable job at modelling a sub-set of behaviours in the real world, as we see here. The CHAIRS simulation as presented is the foundation on which these additional dynamics can be built, but detailed model development need to go hand in hand with enhanced system understanding and with an eye on the policy predictions desired.

Note, also, that though these features could be usefully amended, such changes are contingent on finding the correct individual level datasets, a challenge already identified in the creation of the CHAIRS simulation. Not only is specific data needed for model inputs, complementary data is needed in order to validate the model. Using the ROP dataset illustrated how this process could be challenging. For example, the ROP reports a sizeable number of records for the UK. When this data is disaggregated and examined for smaller districts, the number of records is reduced. For the EASEL district, the final dataset used included just over 600 records. To mitigate this, a process of reweighting was necessary so as to increase the number of records to a population size similar to that of the EASEL district. This reweighting process relied on the 2001 Census, however. The fact remains that barring the 2011 Census which is yet to be released, yearly data containing demographic details as well as households' preferences does not exist. Also, the ROP is limited in reporting a true picture of the distribution of minority groups. This may be due to the reluctance of such groups to engage in survey exercises.

Despite this, the ROP provides a new avenue for collecting research data. It is collected twice a year and includes demographic details as well as household preferences at the individual level. Perhaps this dataset is not yet clearly understood in the academic realm and therefore making this data available for research projects such as this highlights the importance of understanding its spatial and compositional biases. This implies that unless research such as this continues, the full capabilities of this dataset may not be appreciated.

Finally, with regards to computer processing power, the project has been limited due to the extensive processing time required for full model execution. Such a challenge may be mitigated by the use of parallel programming in the supercomputing environment. Also, Repast Simphony provides an ABM tool which includes a module for visualisation. In the context of policy analysis and engaging with policy makers, such a module is attractive in improving the presentation of the results of policy scenarios. However, though this visualisation is possible for small simulations of populations (less than 700 households in the case of the CHAIRS simulation) executing the simulation for the entire EASEL district on a PC was not possible. Though supercomputing facilities were used to mitigate the large run-times (**Section 4.1**), use of the dynamic visualisation module was still not practical. Whether Repast Simphony could be improved to handle larger simulations such as the CHAIRS model is beyond the scope of this research project though dynamic visualisation during model execution could help to reduce the learning curve when the outcomes of this research project are presented to policy stakeholders.

Although the limitations of this research may appear to be substantial, considerable potential has been demonstrated in the simulation approach and if solutions are found to address the challenges identified, research in this field could be further advanced while useful insights could be provided for those engaged in formulating and enacting regeneration policy. ABM is a valuable tool for practical research projects related to human/social geography. In this thesis, it has been used to illustrate the potential outcomes of specific regeneration projects, suggesting that mixed tenure developments can improve socioeconomic diversity across a local community.

Reflecting on the research question and the aims (**Section 1.2**), where housing-led regeneration projects are concerned, the CHAIRS simulation has been created to explore the likely outcomes of at least one regeneration project proposed for the EASEL regeneration district. In this model residential mobility behaviours have been implemented which simulate economic and social drivers and by so doing, various forecasts of the population mix were presented. It may be argued, that some of the behaviours can be further extended, however, the results of the CHAIRS model suggests that housing-led regeneration projects can lead to socially mixed community when these new developments are mixed by housing tenure. Though at the risk of causing gentrified communities, this outcome falls in line with the Leeds City Council's aim of creating more socially mixed communities. Based on the CHAIRS simulation, gentrification could be reduced if the number of low-income housing options is increased in these mixed developments.

Such a contribution is significant to the literature. On one hand, the CHAIRS simulation adds another dimension to the Schelling type models in existence by exploring the potential for increasing the number of behaviours when compared to other residential mobility models. Schelling type models include the work of Laurie *et al.* (2003), Bruch and Mare (2006), Zhang (2004) etc. as discussed in **Section 3.3** and summarised in **Table 3.2**. On the other hand, the CHAIRS simulation is applied to an issue which has been debated in the qualitative literature, that is, whether regeneration projects have positive effects. Using real world data for an existing project engaged in by the Leeds City Council, the model is able to give plausible results. Thus, the CHAIRS simulation is the foundation on which future work can be built from a computer modelling perspective while it adds to the body of applied research and can be extended to other districts undergoing similar regeneration projects in the UK and elsewhere.

List of References

AGUILERA, A. and E. UGALDE. 2007. A Spatially Extended Model for Residential Segregation. *Discrete Dynamics in Nature and Society*. **2007**(Article ID 48589).

ALONSO, W. 1964. *Location and Land Use*. Mass.: Harvard University Press.

AL-RABADI, A. 2011. Conservative Reversible Elementary Cellular Automata and their Quantum Computations. *In*: Alejandro SALCIDO, (ed). *Cellular Automata - Innovative Modelling for Science and Engineering*, Croatia: InTech, pp.57-94.

ATKINSON, R. and K. KINTREA. 2000. Owner occupation, social mix and neighbourhood impacts. *Policy and Politics*. **28**(1), pp.93-108.

AXELROD, R. 2006. Agent-based Modelling as a Bridge Between Disciplines. *In*: Leigh TESFATSION and Kenneth, L JUDD, (eds). *Handbook of Computational Economics*, Iowa: Elsevier/North-Holland, pp.1565-1584.

AXTELL, R. and J. M. EPSTEIN. 1994. Agent-based Modelling: Understanding Our Creations. *The Bulletin of the Santa Fe Institute*. **Winter 1994**, pp.28-32.

BAILEY, T. and A. GATRELL. 1995. *Interactive Spatial Data Analysis*. New York: Prentice Hall.

BALLAS, D and G CLARKE. 2001. Modelling the local impacts of national social policies: a spatial microsimulation approach. *Environment and Planning C: Government Policy*. **19**, pp.587-606.

BALLAS, D., G. P. CLARKE, and I. TURTON. 1999. Exploring microsimulation methodologies for the estimation of household attributes. *In*: *Paper presented at the 4th International Conference on GeoComputation*. Virginia, USA: 25-29 July.

BANKS, J., R. BLUNDELL, Z. OLDFIELD, and J. SMITH. 2011. Housing Mobility and Downsizing at Older Ages in Britain and the United States. *Economica*. **doi:10.1111/j.1468-0335.2011.00878.x**.

BANTON, M. 1994. Modelling ethnic and national relations. *Ethnic and Racial Studies*. **17**(1), pp.1-19.

BEEKMAN, T., F. LYONS, and J. SCOTT. 2001. *Improving the Understanding of the Influence of Owner Occupiers in Mixed Tenure Neighbourhoods*. Edinburgh: Scottish Homes.

BENENSON, I. 2004. Agent-based Modeling: From Individual Residential Choice to Urban Residential Dynamics. *In*: Michael F GOODCHILD and Donald G. JANELLE, (eds). *CSISS Best Practice Publications: Spatially Integrated Social Science*, Oxford: Oxford University Press, pp.67-95.

BENJAMIN, S. C., N. F. JOHNSON, and P. M. HUI. 1996. Cellular automata models of traffic flow along a highway containing a junction. *Journal of Physics A: Mathematics and Theory*. **29**, pp.3119-3127.

BIANCHI, C., P. CIRILLO, M. GALLEGATI, and P. A. VAGLIANSINDI. 2008. Validation in agent-based models: An investigation on the CATS model. *Journal of Economic Behavior and Organization*. **67**(3-4), pp.947-964.

BłASZCZYńSKI, J., K. DEMBCZYńSKI, W. KOTłOWSKI, and M. PAWłOWSKI. 2006. Mining Direct Marketing Data by Ensembles of Weak Learners and Rough Set Methods, Data Warehousing and Knowledge Discovery. *Lecture Notes in Computer Science*. **4081**, pp.218-227.

BOEHM, T. P. 1982. A Hierarchical Model of Housing Choice. *Urban Studies*. **19**(1), pp.17-31.

BÖHEIM, R. and M. TAYLOR. 1999. Residential Mobility, Housing Tenure and the Labour Market in Britain. *Essex University ISER Working Paper*., p.16.

BONABEAU, E. 2002. Agent-based modelling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*. **99**(3), pp.7280-7287.

BRAMLEY, G. and J. MORGAN. 2003. Building Competitiveness and Cohesion: The Role of New Housebuilding in Central Scotland's Cities. *Housing Studies*. **18**(4), pp.447-471.

BRAMLEY, G, H PAWSON, and M MUNRO. 2004. *Key Issues in Housing: Policies and Markets in 21st Century Britain*. Britain: Palgrave Macmillan.

BRATMAN, M., D. ISRAEL, and M. POLLACK. 1988. Plans and Resource-Bounded Practical Reasoning. *Computational Intelligence*. **4**(4), pp.349-355.

BREWER, C. and T. SUCHAN. 2001. *Mapping Census 2000: The Geography of U.S. Diversity*. Washington, DC: US Government Printing Office.

BRIDGE, G. 2002. The Neighbourhood and Social Networks. *Centre for Neighbourhood Research Paper*. **4**(April 2002).

BROWN, L. A. and E. G. MOORE. 1970. The Intra-Urban Migration Process: A Perspective. *Geografiska Annaler. Series B, Human Geography*. **52**(1), pp.1-13.

BRUCH, E. E. and R. D. MARE. 2006. Neighbourhood Choice and Neighbourhood Change. *American Journal of Sociology*. **112**(3), pp.667-709.

BUTLER, T., M. RAMSDEN, and R. WEBBER. 2007. The Best, the Worst, the Average: Secondary School Choice and Education Performance in East London. *Journal of Education Policy*. **22**(1), pp.7-29.

CAGLIONI, M., M. PELIZZONI, and G. A. RABINO. 2006. Urban Sprawl: A Case Study for Project Gigalopolis. *In*: Samira EL YACOUBI, Bastien CHOPARD, and Stefania BANDINI, (eds). *Volume 4173 of Lecture Notes in Computer Science*, Berlin: Springer-Verlag, pp.436-445.

CAMERON, S. 1992. Housing, Gentrification and Urban Regeneration Policies. *Urban Studies*. **29**(1), pp.3-14.

CARLEY, M. and K. KIRK. 1998. *Sustainable by 2020: A strategic approach to urban regeneration for Britains cities*. Bristol: The Policy Press.

CARUSO, G., D. PEETERS, J. CAVAILHÈS, and M ROUNSEVELL. 2007. Spatial configurations in a periurban city. A cellular automata-based microeconomic model. *Regional Science and Urban Economics*. **37**, pp.542-567.

CASTELFRANCHI, C. 1998. Modelling social action for AI agents. *Artificial Intelligence*. **103**(1-2), pp.157-182.

CENTRE FOR CENSUS AND SURVEY RESEARCH. 2011. *The Samples of Anonymised Records*. [online]. [Accessed 26 August 2011]. Available from World Wide Web: <http://www.ccsr.ac.uk/sars/gettingstarted/>

CLARK, W.A.V., M. DEURLOO, and F. DIELEMAN. 2006. Residential Mobility and Neighbourhood Outcomes. *Housing Studies*. **21**(3), pp.323-342.

CLARK, W.A.V. and Y. HUANG. 2003. The life course and residential mobility in British housing markets. *Environment and Planning A*. **35**(2), pp.323-329.

CLARK, W.A.V. and J. L. ONAKA. 1983. Life Cycle and Housing Adjustment as Explanations of Residential Mobility. *Urban Studies*. **20**(1), pp.47-57.

CLARKE, G. 1996. *Microsimulation for Urban and Regional Policy Analysis*. London: Pion Limited.

CLARKE, G., H. EYRE, and C. GUY. 2002. Deriving Indicators of Access to Food Retail Provision in British Cities: Studies of Cardiff, Leeds and Bradford. *Urban Studies*. **39**(11), pp.2041-2060.

COLBURN, T. and G. SHUTE. 2007. Abstraction in Computer Science. *Minds and Machines*. **17**(2), pp.169-184.

CONTOYANNIS, P., A. M. JONES, and N. RICE. 2004. The Dynamics of Health in the British Households Panel Survey. *Journal of Applied Econometrics*. **19**(4), pp.437-503.

CROFT, J. 2004. Positive choice, no choice or total rejection: the perennial problem of school catchments, housing and neighbourhoods. *Housing Studies*. **16**(3), pp.249-265.

CROOKS, A. T. 2010. Constructing and implementing an agent-based model of residential segregation through vector GIS. *International Journal of Geographical Information Science*. **24**(5), pp.661-675.

CROOKS, A, C CASTLE, and M BATTY. 2008. Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*. **32**(6), pp.417-430.

DAVANZO, J. 1981. Repeat Migration, Information Cost, and Location-Specific Capital. *Population and Environment*. **4**(1), pp.45-73.

DAVIS, F. D., R. P. BAGOZZI, and P. R WARSHAW. 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*. **35**(8), pp.982-1003.

DAWKINS, C., T. N. SRINIVASAN, and J. WHALLEY. 2001. Calibration. *In*: J J HECKMAN and E E LEAMER, (eds). *Handbook of Econometrics volume 6*, New York: Elsevier Science, pp.3653-3703.

DE SMITH, M. J., G. F. MICHAEL, and P. A. LONGLEY. 2009. *Geospatial analysis: A comprehensive guide to principles, techniques, and software tools*. Leicester: Matador.

DIELEMAN, F. 2001. Modelling residential mobility; a review of recent rends in research. *Journal of Housing and Built Environment*. **16**(3), pp.249-265.

DIXON, S. 2003. Migration within Britain for Job Reasons. *Labour Market Trends*. **April**, pp.191-201.

DOWNWARD, P. 2007. Exploring the Economic Choice to Participate in Sport: Results from the 2002 General Household Survey. *International Review of Applied Economics*. **21**(5), pp.633-653.

DUTTON, P. 2003. Leeds Calling: The Influence of London and the Gentrification of Regional Cities. *Urban Studies*. **40**(12), pp.2557-2572.

ECONOMIC AND SOCIAL RESEARCH INSTITUTE. 2011. *Understanding Topology and Shapefiles*. [online]. [Accessed 24 August 2011]. Available from World Wide Web: <http://www.esri.com/news/arcuser/0401/topo.html>

FERRARI, E. 2011. Conceptualising Social Housing within the Wider Housing market: A Vacancy Chain Model. *Housing Studies*. **26**(1), pp.95-116.

FORD, J. and R. BURROWS. 1999. The Costs of Unsustainable Home Ownership in Britain. *Journal of Social Policy*. **28**(2), pp.305-330.

FORREST, R. and A. KEARNS. 1999. *Joined-Up Places? Social Cohesion and Neighbourhood Regeneration*. York: YPS/JRF.

FORREST, R and A MURIE. 1988. *Selling the Welfare State: The Privatisation of Public Housing*. London: Routledge.

FORREST, R and A MURIE. 1994. The Dynamics of the Owner-occupied Housing Market in Southern England in the Late 1980s: A Study of New Building and Vacancy Chains. *Regional Studies*. **28**(3), pp.275-289.

FOSSETT, M. and D. R. DIETRICH. 2009. Effects of city size, shape, and form, and neighbourhood size and shape in agent-based models of residential segregation: are Schelling-style preference effects robust? *Environment and Planning B: Planning Design*. **36**(1), pp.149-169.

FOSSETT, M. and R. SENFT. 2004. SIMSEG and Generative Models: A Typology of Model-Generated Segregation Patterns. *In*: *Proceedings of the Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity and Emergence*. Chicago, pp.39-78.

FOSSETT, M and W WAREN. 2005. Overlooked Implications of Ethnic Preferences for Residential Segregation in Agent-based Models. *Urban Studies*. **42**(11), pp.1893-1917.

FOSSETT, M. and W. WARREN. 2005. Overlooked Implications of Ethnic Preferences for Residential Segregation in Agent-based Models. *Urban Studies*. **42**(11), pp.1893-1917.

FOTHERINGHAM, A. and M. O'KELLY. 1989. *Spatial Interaction Models: Formulations and Applications*. Dordecht: Kluwer.

FRANKLIN, S. and A. GRAESSER. 1996. Is it an agent, or just a program? A taxonomy for autonomous agents. *In*: *Proceedings of the Agent Theories, Architectures, and Languages Workshop*, Berlin: Springer-Verlag.

FURTADO, B.A. 2008. Neighbourhoods in Urban Economics: Incorporating Cognitively Perceived Urban Space in Economic Models. *Urban Studies*. **48**(13), pp.2827-2847.

GIBSON, J., T. KOEPSALL, P. DIEHR, and C. HALE. 1999. Increasing Response Rates for Mailed Surveys of Medicaid Clients and Other Low-income Populations. *American Journal of Epidemiology*. **149**(11), pp.1057-1062.

GJESSING, M. 2010. Interview with R. Jordan. 7 December. Leeds.

GOODMAN, J.L. Jr. 1976. Reasons for moves out of and into large cities. *Journal of the American Planning Association*. **45**, pp.407-416.

GOODMAN, J. L. 1981. Information, Uncertainty, and the Microeconomic Model of Migration Decision Making. *In*: Gordon, F DE JOHN and Robert w GARDNER, (eds). *Migration Decision Making: Multidisciplinary Approaches to Microlevel Studies in Developed and Developing Countries*, NY: Pergamon Press, pp.13-58.

GOODMAN, A.C. 1995. A Dynamic Equilibrium Model of Housing Demand and Mobility with Transaction Costs. *Journal of Housing Economics*. **4**, pp.307-327.

HANUSHEK, E.A. and J.M. QUIGLEY. 1978. An explicit model of intra-metropolitan mobility. *Land Economics*. **54**, pp.411-429.

HARRELL, F. E. 2001. *Regression modeling strategies: with applications to linear models, logistic regression and survival analysis*. New York: Springer-Verlag.

HAYNES, K. and S. FORTHERINGHAM. 1984. *Gravity and Spatial Interaction Models*. Beverly Hills: Sage.

HEATH, B. L. and R. HILL. 2010. Some insights into the emergence of agent-based modelling. *Journal of Simulation*. **4**, pp.163-169.

HENLEY, A. 1998. Residential Mobility, Housing Equity and the Labour market. *The Economic Journal*. **108**(447), pp.414-427.

HEPPENSTALL, A., A. J. EVANS, and M. H. BIRKIN. 2005. A Hybrid Multi-Agent/Spatial Interaction Model System for Petrol Price Setting. *Transactions in GIS*. **9**(1), pp.35-51.

HICKMAN, P., D. ROBINSON, R. CASEY et al. 2007. *Understanding Housing Demand; Learning from rising markets in Yorkshire and the Humber*. Coventry: The Joseph Rowntree Foundation.

HILL, R. R. Mcintyre, G. A. Narayanan, S. 2001. Genetic Algorithms for Model Optimization. *In*: *Simulation Technology and Training Conference (SimTechT)*. Australia: Simulation Industry Association of Australia.

HINCKS, S. and C. WONG. 2010. The Spatial Interaction of Housing and Labour Markets: Commuting Flow Analysis of North West England. *Urban Studies*. **47**(3), pp.620-649.

HOLDEN, C. 1996. Cybersocieties: A New Tool for Science. *Science*. **274**, p.727.

HOLMANS, A.E. 1987. *Housing policy in Britain: a history*. New South Wales: Croom Helm Ltd.

HOSMER, D. W. and S. LEMESHOW. 2000. *Applied logistic regression*. Canada: John Wiley & Sons Inc.

HULL, A. 2000. Neighbourhood Renewal: A toolkit for regeneration. *GeoJournal*. **51**(4), pp.301-310.

IHLANDFELDT, K.R. 1981. An empirical investigation of alternative approaches to estimating the equilibrium demand for housing. *Journal of Urban Economics*. **9**, pp.97-105.

Intelligent Agents III. 1997. *In*: J. MUELLER, M. WOOLDRIDGE, and N. R. JENNINGS, (eds). *Lecture Notes on Artificial Intelligence Volume 1193*. Germany: Springer-Verlag, pp.21-36.

JENNINGS, N. R. 2000. On agent-based software engineering. *Artificial Intelligence*. **117**(2), pp.277-296.

JENNINGS, N. R., K. SYCARA, and M. WOOLRIDGE. 1998. A Roadmap of Agent Research and Development. *Autonomous Agent and Multi-Agent Systems*. **1**(1), pp.7-38.

JOHNSTON, R. 2000. Microsimulation. *In*: R J JOHNSTON, Derek GREGORY, Geraldine PRATT, and Michael WATTS, (eds). *The Dictionary of Human Geography*, Oxford: Blackwell.

JOHNSTON, R, S BURGESS, D WILSON, and R HARRIS. 2006. School and Residential Ethnic Segregation: An Analysis of Variations across England's Local Education Authorities. *Regional Studies*. **40**(9), pp.973-990.

JONES, P. 2005. The suburban high flat in the post-war reconstruction of Birmingham, 1945-71. *Urban History*. **32**(2), pp.308-326.

JONES, C. and C. WATKINS. 1996. Urban Regeneration and Sustainable Markets. *Urban Studies*. **33**(7), pp.1129-1140.

JUPP, B. 1999. *Living Together; Community life on mixed tenure estates*. London: Demos.

KEARNS, A. and M. PARKINSON. 2001. The significance of neighbourhood. *Urban Studies*. **38**(12), pp.2103-2110.

KEMP, P. A. and M. KEOGHAN. 2001. Movement Into and Out of the Private Rental Sector in England. *Housing Studies*. **16**(1), pp.21-37.

KIM, J. H., F. PAGLIARA, and J. PRESTON. 2005. The Intention to Move and Residential Location Choice Behaviour. *Urban Studies*. **42**(9), pp.1621-1636.

KLEINHANS, R. 2004. Social implications of housing diversification in urban renewal: a review of recent literature. *Journal of Housing and Built Environment*. **19**(4), pp.367-390.

KLEINMAN, M. 2000. Include Me Out? The New Politics of Place and Poverty. *Policy Studies*. **21**(1), pp.49-61.

KLEINMAN, M. and C.M.E. WHITEHEAD. 1999. Housing and regeneration: the problem or the solution. *National Institute Economic Review*. **October**(170), pp.78-86.

KNUDSEN, D.C. and A.S. FOTHERINGHAM. 1986. Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling. *International Regional Science Review*. **10**(2), pp.127-147.

KONGMUANG, C., G. CLARKE, A. J. EVANS, and D. BALLAS. 2005. Modelling crime victimisation at small-area level using a spatial microsimulation technique. *In*: *Paper presented at the RSAIBIS 35th Annual Conference*. http://www.geog.leeds.ac.uk/people/c.kongmuang/SimCrime%2520Paper_RSAIB%IS.doc.

LAURI, A. J. and N. K. JAGGI. 2003. Role of 'Vision' in Neighbourhood Racial Segregation: A Variant of the Schelling Segregation Model. *Urban Studies*. **40**(13), pp.2687-2704.

LEEDS CITY COUNCIL. 2005. *Enterprise Leeds - Support new business*. [online]. [Accessed 2011 August 24]. Available from World Wide Web: <http://www.leeds.gov.uk/page.aspx?pageidentifier=98ae4c51-3dc8-4ac4-9298-cf3f79b16ed6>

LEEDS CITY COUNCIL. 2007a. *East and South East Leeds Area Action Plan*. Leeds: Leeds City Council Development Department.

LEEDS CITY COUNCIL. 2007b. *East and South East Leeds Housing Needs and Aspirations Study*. Leeds: Outside Research and Development.

LEEDS CITY COUNCIL. 2007c. *Strategic Housing Market Assessment*. Leeds: Outside Research and Development.

LEE, B. A., R. S. OROPESA, and J. W. KANA. 1994. Neighbourhood Context and Residential Mobility. *Demography*. **31**(2), pp.249-270.

LEE, L.F. and R. TROST. 1978. Estimation of some limited dependent variable models with application to housing demand. *Journal of Econometrics*. **8**, pp.357-382.

LONG, L. H. 1972. The Influence of Number and Ages of Children on Residential Mobility. *Demography*. **9**(3), pp.371-382.

LONG, L. 1998. *Migration and Residential Mobility in the United States*. New York: Russell Sage Foundation.

LOUIE, M. A. and K. CARLEY. 2008. Balancing the criticisms: Validating multi-agent models of social systems. *Simulation Modelling Practice and Theory*. **16**(2), pp.242-256.

MACAL, C. M. and M. J. NORTH. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation*. **4**, pp.151-162.

MAGIDSON, J. 1993. *SPSS for Windows, CHAID, release 6.0*. Chicago, Ill.: SPSS Inc.

MAHDAVI, B., D. O'SULLIVAN, and P. DAVIS. 2007. An Agent-based Microsimulation Framework For Investigating Residential Segregation Using Census Data. *In*: *Paper presented at the MODSIM 2007 International Congress on Modelling and Simulation*. Christchurch, New Zealand.

MALLESON, N., A. HEPPENSTALL, and L. SEE. 2008. Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems*. **31**(3), pp.236-250.

MALPASS, P. 1999. Housing Policy: does it have a future? *Policy & Politics*. **27**(2), pp.217-228.

MALPASS, P. 2000. Public utitlity societies and the Housing and Town Planning Act, 1919: a re-examination of the introduction of state-subsidized housing in Britain. *Planning Perspectives*. **15**(4), pp.377-392.

MANN, H. B. and D. R. WHITNEY. 1947. On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*. **18**(1), pp.50-60.

MASSEY, D S. 1996. The Age of Extremes: Concentrated Affluence and Poverty in the Twenty-first Century. *Demography*. **3**(4), pp.395-412.

MASSEY, D. S. and M. J. FISCHER. 2000. How segregation concentrates poverty. *Ethnic and Racial Studies*. **23**(4), pp.670-691.

MCCARTHY, K. 1976. The Household Life Cycle and Housing Choices. *Papers in Regional Science*. **37**(1), pp.55-80.

MCDONALD, S., N. MALYS, and V. MALIENE. 2009. Urban regeneration for sustainable communities: A case study. *Technological and Economic Development of Economy*. **15**(1), pp.49-59.

MCGREGOR, A. and D. MACLENNAN. 1992. *A review and critical evaluation of strategic approaches to urban regeneration*. Edinburgh: Scottish Homes.

MEGBOLUGBE, I. F. and P. D. LINNEMAN. 1993. Home Ownership. *Urban Studies*. **30**(4/5), pp.659-682.

MERRETT, S. 1979. *State Housing in Britain*. Henley-on-Thames: Routledge & Kegan Paul.

MILLER, M. J. 1959. *New Life for Cities Around the World. International Handbook on Urban Renewal*. New York: Books International.

MÓRA, M. C., J. G. LOPESY, R. M. VICCARIZ, and C. HELDER. 1999. BDI Models and Systems: Reducing the Gap. *In*: *Proceedings of the 5th International Workshop on Intelligent Agents V, Agent Theories, Architectures, and Languages, LNAI 1555*. London: Springer-Verlag, pp.11-27.

MULDER, C. H. 1996. Housing Choice: Assumptions and Approaches. *Journal of Housing and Built Environment*. **11**(3), pp.209-232.

MULDER, C. H. and P. HOOIMEIJER. 1999. Residential relocations in the life course. *In*: Leo J G VAN WISSEN and Pearl A DYKSTRA, (eds). *Population Issues: an interdisciplinary focus*, New York: Kluwer-Academic/Plenum Publishers, pp.159-186.

MULLINS, D., A. MURIE, P. LEATHER et al. 2006. *Housing Policy in the UK*. Hampshire: Palgrave Macmillan.

NIAZI, M. and A. HUSSAIN. 2009. Agent-Based Tools for Modeling and Simulation of Self-Organization in Peer-toPeer, Ad Hoc, and Other, Complex Networks. *IEEE Communications Magazine*. **47**(3), pp.166 - 173.

OFFICE OF NATIONAL STATISTICS. 2011a. *Census Geography*. [online]. [Accessed 2011 August 24]. Available from World Wide Web: <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census.index.html>

OFFICE OF NATIONAL STATISTICS. 2011b. *Information Page*. [online]. [Accessed 2011 Sep 8]. Available from World Wide Web: <http://www.neighbourhood.statistics.gov.uk/dissemination/Info.do?page=userguide/detailedguidance/casestudies/geography/case-studies-geography.htm>

O'SULLIVAN, D. 2009. Changing neighborhoods - neighborhoods changings: a framework for spatially explicit agent-based models of social systems. *Sociological Methods and Research*. **37**(4), pp.493-530.

O'SULLIVAN, D., J. MACGILL, and C. YU. 2003. Agent-Based Residential Segregation: A Hierarchically Structured Spatial Model. *In*: MACAL, NORTH, and SALLACH, (eds). *Agent 2003: Challenges in Social Simulation*, Chicago: Argonne National Laboratory, pp.493-507.

PALMER, G., T. MCINNES, and P. KENWAY. 2008. *Monitoring poverty and social exclusion*. York: The Joseph Rowntree Foundation.

PANCS, R. and N. J. VRIEND. 2007. Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics*. **91**(1-2), pp.1-24.

PANZARASA, P., T. NORMAN, and N. JENNINGS. 2001. Going Public and the Sale of Shares with Heterogeneous Investors: Agent-Based Computational Modelling and Computer Simulations. *Group Decisions and Negotiation*. **10**(5), pp.423-470.

PARKER, D. 2006. Modelling land-use and land-cover change. *In*: *Land-use and Land-cover Change: Local Process, Global Impacts*, New York: Berlin Heidelberg.

PAWSON, H. and G. BRAMLEY. 2004. *Key Issues in Housing: Markets and Policies in 21st Century Britain*. Palgrave Macmillan.

PEACH, C. 1996. Does Britain Have Ghettos? *Transactions of the Institute of British Geographers*. **21**(1), pp.216-235.

PHILLIPS, D. 1981. The social and spatial segregation of Asians in Leicester. *In*: Peter JACKSON and Susan J SMITH, (eds). *Social Interaction and Ethnic Segregation*, London: Academic Press, pp.101-121.

PHILLIPS, D. 1988. Race and housing in London's East End: continuity and change. *New Community*. **14**(3), pp.356-369.

PHILLIPS, D. 1998. Black Minority Ethnic Concentration, Segregation and Dispersal in Britain. *Urban Studies*. **35**(10), pp.1681-1702.

PHILLIPS, D, C DAVIS, and F BUTT. 2002. The Racialisation of Space in Bradford. *The Yorkshire and Humber Regional Review*. **12**(2), pp.9-10.

POWER, A. 1993. *Hovels to High Rises State Housing in Europe since 1850*. London: Routledge.

POWER, A. & THE JOSEPH ROWNTREE FOUNDATION. 1999. *The Slow Death of Great Cities? Urban Abandonment or Urban Renaissance*. York: YPS.

QUIGLEY, J.M. and D.H. WEINBERG. 1977. Intraurban residential mobility: a review and synthesis. *International Regional Science Review*. **2**, pp.41-66.

RABE, B. and M. P. TAYLOR. 2010. Residential mobility, quality of neighbourhood and life course events. *Journal of the Royal Statistical Society*. **173**(3), pp.532-555.

RAO, A. and M. GEORGEFF. 1995. BDI Agents: From Theory to Practice. *In*: *Proceedings of the First International Conference on Multiagent Systems*, San Francisco: AAAI Press, pp.312-319.

REES, P., J. STILLWELL, and A. TYLER-JONES. 2004. The city is the people: demographic structure and dynamics. *In*: Unsworth R and J STILLWELL, (eds). *21st century Leeds: geographies of a regional city*, Leeds: Leeds University Press, pp.26-48.

REES, P., A. TYLER-JONES, and J.C.H STILLWELL. 2004. Population structure and life expectancy: community area dynamics in Leeds, 1991-2001. *In*: RSG (with IBG) Population Geography Research GROUP, (ed). *'Population Geographies'*. University of St. Andrews.

RESNICK, M. 1997. *Turtles, termites, and traffic jams. Explorations in massively parallel microworlds*. Cambridge: MIT Press.

ROSSI, P. H. 1955. *Why families move: a study in the social psychology of urban residential mobility*. Glencoe, IL: Free Press.

RUSSELL, S. and P. NORVIG. 2003. *Artificial Intelligence A Modern Approach*. New Jersey: Pearson Education Inc.

SARGENT, T. J. 1998. Verification and validation in simulation models. *In*: S G HENDERSON, B BILLER, M H HSIEH et al., (eds). *Proceedings of the Winter Simulation Conference*., pp.52-64.

SCHELLING, T. C. 1969. Models of Segregation. *The American Economic Review*. **59**(2), pp.488-493.

SCHELLING, T. C. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology*. **1**(2), pp.143-186.

SCHMIDT, B. 2000. *The Modelling of Human Behaviour*. Erlangen: SCS Publications.

SCHOOL ACCESS SERVICES. 2011. *Policy for the provision of home to school or college transport for children and students prior to their nineteenth birthday*. Leeds City Council: Leeds.

SENIOR, M, C.J. WEBSTER, and N BLANK. 2006. Residential relocation and sustainable urban form: statistical analysis of owner-occupiers' preferences. *International Planning Studies*. **11**(1), pp.41-57.

SEO, J. K. 2002. Re-urbanisation in Regenerated Areas of Manchester and Glasgow, New Residents and the Problems of Sustainability. *Cities*. **19**(2), pp.113-121.

SETHI, R. and R. SOMANATHAN. 2004. Inequality and Segregation. *Journal of Political Economy*. **112**(6), pp.1296-1321.

SHOOMAN, M. L. 2002. *Reliability of Computer Systems and Networks: Fault Tolerance, Analysis, and Design*. New York: John Wiley & Sons Inc.

SMITH, D., G. CLARKE, and K. HARLAND. 2009. Improving the synthetic data generation process in spatial microsimulation models. *Environment and Planning A*. **41**, pp.1251-1268.

SOUTH, S. J. and K. D. CROWDER. 1997. Residential Mobility Between Cities and Suburbs: Race, Suburbanization, and Back-to-the-City Moves. *Demography*. **34**(4), pp.525-538.

STERMAN, J. D. 1988. A Skeptic's Guide to Computer Models. *In*: G P RICHARDSON, (ed). *Modelling for Management*, Aldershot, UK,: Dartmouth Publishing Company, pp.133-169.

STEWART, J. and M. RHODEN. 2003. A review of social housing regeneration in the London Borough of Brent. *The Journal of the Royal Society for the Promotion of Health*. **March 123**(1), pp.23-32.

STILLWELL, J., S. HUSSAIN, and P. NORMAN. 2008. The internal migration propensities and net migration patterns of ethnic groups in Britain. *Migration Letters*. **5**(2), pp.135-150.

STILLWELL, J. and D. PHILLIPS. 2006. Diversity and Change: Understanding the Ethnic Geographies of Leeds. *Journal of Ethnic and Migration Studies*. **7**(32), pp.1131-1152.

STRASSMAN, W. P. 2001. Residential Mobility: Contrasting Approaches in Europe and the United States. *Housing Studies*. **16**(1), pp.7-20.

THOMPSON, C., J. STILLWELL, M. CLARK, and C. BRADBROOK. 2010. Understanding and Validating Acxiom's Research Opinion Poll Data. *Working Paper Series 10/07, School of Geography, University of Leeds*.

TIMMS, D. W.G. 1971. The Urban Mosaic: Towards a Theory of Residential Differentiation. *The Town Planning Review*. **42**(4), pp.407-409.

TRIOLA, M. F., W. M. GOODMAN, and R. LAW. 2007. *Elementary Statistics*. Reading: Addison-Wesley Publishing.

TRUEMAN, M., M. KLEMM, and A. GIROUD. 2004. Can a city communicate? Bradford as a corporate brand. *Corporate Communitications: An International Journal*. **9**(4), pp.317-330.

TU, Y. and J. GOLDFINCH. 1996. A Two-stage Housing Choice Forecasting Model. *Urban Studies*. **33**(3), pp.517-537.

TUNSTALL, R. 2003. Mixed tenure policy in the UK: privatisation, pluralism or euphemism? *Housing, Theory and Society*. **20**(3), pp.153-159.

TUROK, I. 1992. Property-led urban regeneration: panacea or placebo? *Environment and Planning A*. **24**, pp.361-379.

UITERMARK, J. 2003. "Social mixing" and the management of disadvantaged neighbourhoods: The Dutch policy of urban restructuring revisited. *Urban Studies*. **40**(3), pp.531-549.

VICKERS, D. and P. REES. 2007. Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society, Series A*. **170**(2), pp.379-403.

WEBSTER, F. V., BLY, P. H., and PAULLEY, N. J. (eds). 1988. Urban land-use and transport interaction: Policies and models. Berkshire, UK: Avebury, Aldershot.

WHITEHEAD, C. 1993. Privatizing Housing: An Assessment of UK Experience. *Housing Policy Debate*. **4**(1), pp.101-139.

WILLIAMSON, P., M. H. BIRKIN, and P. REES. 1998. The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A*. **30**(5), pp.785-816.

WILSON, W. J. 1987. *The Truly Disadvantaged: The Inner City, the Underclass and Public Policy*. Illinois: Unversity of Chicago Press.

WOHLAND, P, P REES, P NORMAN et al. 2010. Ethnic population projections for the UK and local areas, 2001-2051. *School of Geography, University of Leeds*. **WP 10/02**.

WOOD, G., R. WATSON, and P. FLATAU. 2006. Microsimulation Modelling of Tenure Choice and Grants to Promote Home Ownership. *The Australian Economic Review*. **39**(1), pp.14-34.

YIN, L. 2009. The Dynamics of Residential Segregation in Buffalo: An Agent-based Simulation. *Urban Studies*. **46**(13), pp.2749-2770.

ZERGER, A. and S. WEALANDS. 2004. Beyond Modelling: Linking Models with GIS for Flood Risk Management. *Natural Hazards*. **33**(2), pp.191-208.

ZHANG, J. 2004. A Dynamic Model of Residential Segregation. *The Journal of Mathematical Sociology*. **28**(3), pp.147-170.

List of Abbreviations

AAP Area Action Plan

ABM Agent-Based Modelling/Models

ARC1 Advanced Research Computing (1st phase)

BDI Beliefs, Desires, Intentions

BHPS British Household Panel Survey

BME Black and Minority Ethnic groups

CA Cellular Automaton

CBD Central Business District

CHAID Chi-Squared Automatic Interaction Detector

CHAIRS Creating Housing Alternatives in Regenerated Societies

CCSR Centre for Census and Survey Research

DI Diversity Index

EASEL East and South East Leeds

EFS Expenditure and Food Survey

EHS English Housing Survey

ESRI Economic Social Research Institute

FMF Flexible Modelling Framework

GHS General Household Survey

IDE Integrated Development Environment

IMD Index of Multiple Deprivation

IoS Index of Segregation

JRF Joseph Rowntree Foundation

LFS Labour Force Survey

LLSOA Lower Level Super Output Area

MLSOA Mid-Level Super Output Area

NDC New Deal for Communities

NGS National Grid Service

OA Output Area

PECS Physical Conditions, Emotional State, Cognitive Capabilities and Social Status

PLASC Pupil Level Annual School Census

ROP Research Opinion Poll

RSL Registered Social Landlords

SAE Standardised Absolute Error

SAR Sample of Anonymised Records

SAS Small Area Statistics

SEH Survey for English Housing

SRMSE Square Root of the Mean Square Error

TAE Total Absolute Error

Appendices

A Household SAR variables defined

Accommodation Type
1 = Detached
2 = Semi-detached
3 = Terraced house / bungalow
4 = Purpose built flats
5 = Flat converted or shared house (including bed-sits)
6 = Flat, maisonette in commercial building
7 = Mobile or temporary structure

Age
1 = 16-25
2 = 26-35
3 = 36-45
4 = 46-55
5 = 56-65
6 = 66-75
7 = > 75

Car Ownership
0 = No car
1 = 1 car
2 = 2 or more cars

Country of Birth
1 = England
2 = Scotland
3 = Republic of Ireland
4 = Northern Ireland
5 = Wales
6 = UK (England, Wales, Scotland)
7 = Ireland, part no specified (NI only)
8 = Western Europe
9 = Eastern Europe
10 = India
11 = Pakistan and Bangladesh
12 = Rest of Asia
13 = Caribbean
14 = North America
15 = Africa
16 = Other

Density
1 = Up to 0.5
2 = Over 0.5 and up to 0.75
3 = Over 0.75 and up to 1
4 = Over 1 and up to 1.5
5 = Over 1.5

Number of Persons in Household
1 – 12+

Economic Activity
1 = In employment (employee or self-employed)
2 = Unemployed
3 = Student not economically active
4 = Other economically inactive

Ethnicity
1 = White British
2 = White Irish
3 = Other White
4 = Mixed White and Black Caribbean
5 = Mixed White and Black African
6 = Mixed White and Asian
7 = Other Mixed
8 = Indian
9 = Pakistani
10 = Bangladeshi
11 = Other Asian
12 = Black Caribbean
13 = Black African
14 = Other Black
15 = Chinese
16 = Other ethnic group

Family Type
1 = Lone parent
2 = Married couple, no children
3 = Married couple with children
4 = Cohabiting couple, no children
5 = Cohabiting couple with children
6 = Ungrouped individual

Housing Tenure
1 = Owns outright/without mortgage/shared ownership
2 = Rents from Council / Local Authority or rents from Housing Association
3 = Private rented or lives rent free

Social Class
1 = Large organisation, Higher Managers and Professionals
2 = Lower Managerial
3 = Intermediate Occupations
4 = Small employers and own account workers
5 = Lower Supervisory
6 = Semi-routine occupations
7 = Routine occupations
8 = Never worked and long-term unemployed
9 = Not classified (including full time students)

Propensity to Move
0 = Not moved
1 = Moved

Number of rooms required
0 – 30

Number of rooms in occupied house
01 – 15+

B Decision Tree defined

<table>
<tr><th>Age</th><th>Tenure</th><th>Family Type</th><th>Number of residents in house</th></tr>
<tr><td rowspan="5">16-25</td><td rowspan="3">Private Renters</td><td rowspan="2">Married Couple no children, Cohabiting couple no children, Ungrouped Individuals</td><td><=2 (1.13%)</td></tr>
<tr><td>> 2 (0.47%)</td></tr>
<tr><td>Lone parent, Married couple with children, Cohabiting couple with children (0.37%)</td><td></td></tr>
<tr><td rowspan="2">Owners, Social Housing Tenants</td><td>Cohabiting couple no children, Cohabiting couple with children, Ungrouped individuals (1.48%)</td><td rowspan="2"></td></tr>
<tr><td>Lone parent, Married couple no children, Married couple with children, Cohabiting couple with children (1.09%)</td></tr>
</table>

16-25 Decision Tree Branch

<table>
<tr><th>Age</th><th>Tenure</th><th>Family Type</th><th>Number of residents in house/Ethnicity</th><th>Qualifications</th></tr>
<tr><td rowspan="8">26-35</td><td rowspan="3">Private Renters</td><td rowspan="2">Married couple no children, Cohabiting no children, Cohabiting with children, Ungrouped individual</td><td>Number of residents in house <= 1 (1.12%)</td><td rowspan="3"></td></tr>
<tr><td>Number of residents in house > 1 (1.38%)</td></tr>
<tr><td>Lone parent, Married couple no children, Married couple with children, Cohabiting couple with children (1.05%)</td><td></td></tr>
<tr><td rowspan="5">Owners, Social Housing Tenants</td><td rowspan="3">Lone parent, Married couple no children, Married couple with children, Cohabiting couple with children, Ungrouped individual</td><td rowspan="2">White British, White Irish, Bangladeshi, Black Caribbean, Other Black, Chinese</td><td>No qualifications, Level 1, Level 2, Level 3, Other qualifications (5.55%)</td></tr>
<tr><td>Level 4/5 (1.81%)</td></tr>
<tr><td>White Other, Mixed Black and White Caribbean, Mixed Black and White African, Mixed White and Asian, Mixed Other, Indian, Pakistani, Other Asian, Black African, Other Ethnic Group (0.76%)</td><td></td></tr>
<tr><td>Cohabiting couple no children, Cohabiting couple with children (1%)</td><td colspan="2" rowspan="2"></td></tr>
<tr><td>Married couple with children (3.7%)</td></tr>
</table>

26-35 Decision Tree Branch

<table>
<tr><th>Age</th><th>Tenure</th><th>Social Class/Family Type</th><th>Number of Rooms in house</th></tr>
<tr><td rowspan="7">36-45</td><td rowspan="2">Private Renters</td><td>Higher Managerial, Lower Managerial, Intermediate Occupations, Small employers, Never worked, No Occupation code (1.17%)</td><td rowspan="5"></td></tr>
<tr><td>Lower Supervisory, Semi-routine jobs, Routine jobs, Long-term unemployed, Full-time students, Over 75 (0.95%)</td></tr>
<tr><td>Social Housing Tenants (3.51%)</td><td></td></tr>
<tr><td rowspan="4">Owners</td><td>Lone parent, Married no children, Cohabiting with children, Ungrouped individuals (6.18%)</td></tr>
<tr><td>Cohabiting couple no children (1.14%)</td></tr>
<tr><td rowspan="2">Married with children</td><td><= 8 (6.27%)</td></tr>
<tr><td>> 8 (0.94%)</td></tr>
</table>

36-45 Decision Tree Branch

<table>
<tr><th>Age</th><th>Tenure</th><th>Family Type</th><th>Rooms Required</th><th>Accommodation Type/Family Type</th></tr>
<tr><td rowspan="7">46-55</td><td>Private Renters (1.41%)</td><td rowspan="5">Married no children, Married with children, Ungrouped individuals</td><td rowspan="3"><= 3</td><td rowspan="2">Purpose-built Flat, Shared house, Maisonette, Temporary structure (1.28%)</td></tr>
<tr><td rowspan="6">Owners, Social Housing Tenants</td></tr>
<tr><td>Detached, Semidetached, Terrace (5.63%)</td></tr>
<tr><td rowspan="2">> 3</td><td>Married no children, Ungrouped individuals (0.47%)</td></tr>
<tr><td>Married with children (6.42%)</td></tr>
<tr><td>Lone parent, Married with children, Cohabiting no children (3.09%)</td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>Cohabiting with children (0.31%)</td></tr>
</table>

46-55 Decision Tree Branch

<table>
<tr><th>Age</th><th>Rooms Required</th><th>Tenure/Family Type</th><th>Accommodation Type</th></tr>
<tr><td rowspan="5">56-65</td><td rowspan="3"><= 3</td><td>Social Housing Tenants, Private Renters (2.58%)</td><td></td></tr>
<tr><td rowspan="2">Owners</td><td>Purpose-built Flat, Shared house, Maisonette, Temporary structure (0.56%)</td></tr>
<tr><td>Detached, Semidetached, Terrace (7.72%)</td></tr>
<tr><td rowspan="2">> 3</td><td>Lone parent, Married with children, Cohabiting with children, Ungrouped individuals (3.29%)</td><td rowspan="2"></td></tr>
<tr><td>Married no children, Cohabiting no children, (0.75%)</td></tr>
</table>

56-65 Decision Tree Branch

<table>
<tr><th>Age</th><th>Accommodation Type</th><th colspan="2">Rooms Required/Rooms occupied in house</th></tr>
<tr><td rowspan="4">>66</td><td>Purpose-built Flat, Shared house, Maisonette, Temporary structure (5.02%)</td><td colspan="2"></td></tr>
<tr><td rowspan="3">Detached, Semidetached, Terrace</td><td rowspan="2">Rooms Required <= 3</td><td>Number of rooms in occupied house <= 4 (4.09%)</td></tr>
<tr><td>Number of rooms in occupied house > 4 (11.68%)</td></tr>
<tr><td>Rooms Required > 3 (2.81%)</td><td></td></tr>
</table>

> 66 Decision Tree Branch

C List of all rule-set combinations

Rule-set	Ethnicity	Social Status	Transport	Schools	Known Areas	Output Areas	Rooms
61	x	x	x				
81	x	x	x	x			
58		x	x	x			
42		x	x				x
52		x	x			x	
91		x	x	x			x
11		x					x
20		x		x			
54	x	x		x			
27		x	x				
82	x	x	x			x	
115	x	x	x	x		x	
18			x			x	
103		x	x	x		x	x
110	x	x	x	x			x
33			x			x	x
69	x		x			x	x
38	x	x					x
72	x	x		x			x
35		x		x			x
5		x					
8						x	x
10	x						x
80	x		x	x		x	
56	x		x	x			
48	x	x				x	
68	x	x	x				x
1							x
83	x	x		x		x	
101	x		x	x		x	x
106	x	x	x			x	x

Rule-set	Ethnicity	Social Status	Transport	Schools	Known Areas	Output Areas	Rooms
92		x	x			x	x
2						**x**	
90				x		x	x
47			x	x		x	
50	x		x			x	
29				x		x	x
23	x	x					
31		x				x	x
71	x			x		x	x
9				x			x
30	x					x	x
19	x			x			
99	x	x		x		x	x
37			x	x			x
16		x				x	
34	x			x			x
14				x		x	
73	x	x				x	x
4	x						
7			x				
121	x	x	x	x		x	x
44	x			x		x	
13			x				x
25	x		x				
3				x			
70	x		x	x			x
45		x		x		x	
15	x					x	
93		x		x		x	x
97		x		x	x	x	
22			x	x			
40	x		x				x
88		x		x	x		x
98		x	x	x		x	
41		x			x		x
113	x	x	x		x		x

Rule-set	Ethnicity	Social Status	Transport	Schools	Known Areas	Output Areas	Rooms
112		x	x	x	x		x
57		x		x	x		
108		x	x		x	x	x
26		x			x		
63		x	x		x		
79	x	x			x	x	
125	x	x	x	x	x		x
102		x		x	x	x	x
126	x	x	x	x	x	x	
67	x	x			x		x
124	x	x	x		x	x	x
60	x	x			x		
95		x	x	x	x		
123		x	x	x	x	x	x
117		x	x	x	x	x	
78	x	x		x	x		
118	x	x	x		x	x	
105	x	x			x	x	x
76	x	x	x		x		
86		x	x		x		x
119	x	x	x	x	x		
96		x	x		x	x	
114	x			x	x	x	
85			x		x	x	x
43			x		x		x
51		x			x	x	
89		x			x	x	x
127	x	x	x	x	x	x	x
64	x		x		x		x
109	x	x		x	x		x
122	x		x	x	x	x	x
107	x		x		x	x	x
87				x	x	x	x
55	x			x	x		
53			x		x	x	
6					x		

Rule-set	Ethnicity	Social Status	Transport	Schools	Known Areas	Output Areas	Rooms
65	x			x	x		x
120	x	x		x	x	x	x
12					x		x
49	x				x	x	
21				x	x		
75	x		x		x	x	
17					x	x	
116	x	x	x	x	x	x	
66	x				x	x	x
32					x	x	x
111	x		x	x	x		x
24	x				x		
39	x				x		x
104			x	x	x	x	x
36				x	x		x
84			x	x	x		x
28			x		x		
100	x			x	x	x	x
46				x	x	x	
59			x	x	x		
62	x		x		x		
74	x		x	x	x		
77	x			x	x	x	
94		x	x	x	x		

D Results for all rule-sets using the Standardised Absolute Error

Rule-set	Tenure	Accommodation Type	Ethnicity	Age	Average SAE	Rule-set Rank
61	0.320578	1.067922306	0.251202	0.616413	0.564029	1
81	0.326088	1.073338385	0.24509	0.614168	0.564671	2
58	0.315919	1.076712816	0.254904	0.612342	0.564969	3
42	0.322209	1.069648769	0.250162	0.617976	0.564999	4
52	0.348378	1.059714035	0.24818	0.609381	0.566413	5
91	0.33511	1.0587134	0.253872	0.618896	0.566648	6
11	0.336915	1.053940777	0.247505	0.631548	0.567477	7
20	0.333974	1.062746841	0.260375	0.614935	0.568008	8
54	0.332045	1.057212391	0.25928	0.626189	0.568682	9
27	0.325022	1.077858397	0.247308	0.625119	0.568827	10
82	0.35964	1.055531835	0.240645	0.61963	0.568861	11
115	0.357838	1.056167787	0.244562	0.617685	0.569063	12
18	0.356681	1.055889733	0.247216	0.616734	0.56913	13
103	0.350392	1.066018387	0.245643	0.614658	0.569178	14
110	0.323374	1.083289647	0.253412	0.616946	0.569255	15
33	0.357252	1.054962953	0.247443	0.618351	0.569502	16
69	0.369055	1.056111302	0.242121	0.611657	0.569736	17
38	0.338772	1.056173726	0.249024	0.635006	0.569744	18
72	0.335346	1.06348136	0.254511	0.626808	0.570037	19
35	0.338933	1.063780802	0.249746	0.628593	0.570263	20
5	0.332011	1.057996431	0.255317	0.638706	0.571007	21
8	0.38118	1.04294874	0.241795	0.618546	0.571118	22
10	0.386098	1.032428133	0.248265	0.6211	0.571973	23
80	0.368418	1.045316049	0.251247	0.624962	0.572486	24
56	0.386289	1.048739178	0.253571	0.605958	0.573639	25
48	0.367221	1.054184477	0.255535	0.617761	0.573675	26
68	0.342537	1.075041137	0.251487	0.625979	0.573761	27
1	0.365725	1.043508442	0.252154	0.63556	0.574237	28
83	0.36003	1.05864821	0.251558	0.627438	0.574419	29
101	0.382895	1.05880132	0.246836	0.609327	0.574465	30
106	0.345663	1.075745045	0.255573	0.621752	0.574683	31
92	0.358954	1.075266648	0.239138	0.627478	0.575209	32

Rule-set	Tenure	Accommodation Type	Ethnicity	Age	Average SAE	Rule-set Rank
2	0.375245	1.049754074	0.243714	0.633401	0.575528	33
90	0.376869	1.070188517	0.243389	0.612105	0.575638	34
47	0.383747	1.056381798	0.248123	0.61539	0.57591	35
50	0.379332	1.059819211	0.242609	0.622161	0.57598	36
29	0.385642	1.050950847	0.250113	0.617829	0.576134	37
23	0.335137	1.080173953	0.255371	0.634278	0.57624	38
31	0.36703	1.068386644	0.244336	0.625956	0.576427	39
71	0.398453	1.044946298	0.25121	0.611177	0.576447	40
9	0.370469	1.048803866	0.259169	0.627689	0.576533	41
30	0.381694	1.048934846	0.243713	0.632037	0.576595	42
19	0.3812	1.042277382	0.258594	0.624672	0.576686	43
99	0.368085	1.053751142	0.261593	0.624552	0.576996	44
37	0.392864	1.054311212	0.254905	0.606852	0.577233	45
16	0.368358	1.062552358	0.249415	0.629185	0.577378	46
34	0.39091	1.05097717	0.25222	0.616256	0.577591	47
14	0.393053	1.044540933	0.255365	0.618566	0.577881	48
73	0.372562	1.055733957	0.241426	0.641971	0.577923	49
4	0.386299	1.040579374	0.251614	0.633427	0.57798	50
7	0.3868	1.061038822	0.252003	0.614344	0.578546	51
121	0.357572	1.07800536	0.258544	0.621253	0.578844	52
44	0.399317	1.042083789	0.253975	0.62078	0.579039	53
13	0.388233	1.060281381	0.248306	0.620418	0.579309	54
25	0.381944	1.062289179	0.255358	0.619927	0.57988	55
3	0.390697	1.056037431	0.2607	0.612235	0.579917	56
70	0.399466	1.056264387	0.248025	0.618288	0.580511	57
45	0.369115	1.06978327	0.248535	0.636629	0.581016	58
15	0.389836	1.0463222	0.2544	0.637117	0.581919	59
93	0.365204	1.078610895	0.248566	0.636359	0.582185	60
97	0.381674	1.11331638	0.247188	0.58914	0.58283	61
22	0.386967	1.066258728	0.255672	0.623009	0.582977	62
40	0.389606	1.057253194	0.255817	0.631698	0.583594	63
88	0.337313	1.128942676	0.243647	0.630736	0.58516	64
98	0.357995	1.074706178	0.249756	0.662355	0.586203	65
41	0.336316	1.13462809	0.245466	0.633757	0.587542	66
113	0.343298	1.148313206	0.23725	0.623935	0.588199	67

Rule-set	Tenure	Accommodation Type	Ethnicity	Age	Average SAE	Rule-set Rank
112	0.339764	1.141444842	0.24569	0.626962	0.588465	68
57	0.335427	1.136222298	0.245407	0.640029	0.589271	69
108	0.368605	1.128352219	0.231396	0.629417	0.589442	70
26	0.329878	1.151016158	0.235306	0.641641	0.58946	71
63	0.335753	1.157901805	0.235193	0.630297	0.589786	72
79	0.375992	1.119928002	0.234706	0.631343	0.590492	73
125	0.337664	1.152669033	0.250401	0.624034	0.591192	74
102	0.37071	1.11821303	0.243384	0.632717	0.591256	75
126	0.369273	1.121781348	0.242348	0.632241	0.591411	76
67	0.329059	1.142712623	0.248623	0.646737	0.591783	77
124	0.368285	1.134880746	0.241248	0.624522	0.592234	78
60	0.329562	1.147777166	0.24935	0.64323	0.59248	79
95	0.337408	1.156044519	0.250466	0.626152	0.592517	80
123	0.36879	1.129499984	0.24813	0.623918	0.592585	81
117	0.368099	1.136004976	0.242157	0.624246	0.592627	82
78	0.335716	1.13412835	0.248291	0.653344	0.59287	83
118	0.375309	1.119379691	0.240038	0.637019	0.592936	84
105	0.379238	1.12005831	0.248082	0.625849	0.593307	85
76	0.339303	1.159528245	0.243087	0.634879	0.594199	86
86	0.342686	1.148712913	0.246613	0.641811	0.594956	87
119	0.331147	1.157404098	0.24188	0.649726	0.595039	88
96	0.366613	1.131995082	0.238502	0.644386	0.595374	89
114	0.36812	1.118621884	0.250636	0.650417	0.596949	90
85	0.394708	1.122593774	0.236726	0.633876	0.596976	91
43	0.393475	1.128809195	0.240437	0.626402	0.597281	92
51	0.368676	1.129216864	0.244969	0.646456	0.597329	93
89	0.369659	1.132576782	0.248648	0.639585	0.597617	94
127	0.371696	1.131228899	0.243841	0.645184	0.597987	95
64	0.404175	1.126314991	0.246705	0.618336	0.598883	96
109	0.341259	1.153382105	0.249934	0.651113	0.598922	97
122	0.415341	1.122635205	0.233685	0.625256	0.599229	98
107	0.398368	1.133217887	0.243761	0.623642	0.599747	99
87	0.411438	1.110986979	0.249444	0.627324	0.599798	100
55	0.393002	1.124494334	0.24566	0.636802	0.59999	101
53	0.404616	1.124967557	0.235602	0.635043	0.600057	102
6	0.388143	1.128372693	0.248613	0.637789	0.600729	103

Rule-set	Tenure	Accommodation Type	Ethnicity	Age	Average SAE	Rule-set Rank
65	0.382489	1.122180536	0.250928	0.648498	0.601024	104
120	0.378714	1.124588416	0.251879	0.649353	0.601134	105
12	0.389538	1.130684522	0.249108	0.636292	0.601406	106
49	0.410841	1.122557254	0.244336	0.628243	0.601494	107
21	0.388269	1.127639579	0.253795	0.638547	0.602062	108
75	0.407715	1.137335814	0.236869	0.630688	0.603152	109
17	0.414353	1.118432721	0.236938	0.64423	0.603488	110
116	0.417957	1.126363112	0.240665	0.629031	0.603504	111
66	0.411137	1.124156509	0.2382	0.64058	0.603518	112
32	0.408963	1.127624918	0.237543	0.640504	0.603659	113
111	0.383708	1.139541909	0.252752	0.639282	0.603821	114
24	0.400234	1.128086808	0.246229	0.643109	0.604415	115
39	0.391945	1.132231696	0.247203	0.648245	0.604906	116
104	0.410689	1.130158392	0.246497	0.635234	0.605645	117
36	0.399767	1.133842398	0.246461	0.645007	0.606269	118
84	0.414446	1.131632615	0.249297	0.630911	0.606572	119
28	0.411294	1.135570662	0.246152	0.633298	0.606579	120
100	0.420918	1.113399098	0.236881	0.655243	0.60661	121
46	0.415801	1.1205296	0.248714	0.642699	0.606936	122
59	0.408886	1.142811159	0.244444	0.635408	0.607887	123
62	0.403147	1.147915923	0.244855	0.635644	0.607891	124
74	0.404235	1.144661546	0.247596	0.63516	0.607913	125
77	0.41254	1.12877641	0.244027	0.648712	0.608514	126
94	0.422539	1.141902035	0.248494	0.658855	0.617947	127

E Results for all rule-sets using the Total Absolute Error

Rule-set	Tenure	Accommodation Type	Ethnicity	Age
61	16.02889	53.3961153	12.56011	30.82065
81	16.3044	53.66691924	12.25448	30.70838
58	15.79596	53.83564078	12.7452	30.61709
42	16.11045	53.48243846	12.50808	30.89882
52	17.41892	52.98570176	12.40901	30.46904
91	16.75549	52.93566999	12.69361	30.94478
11	16.84577	52.69703887	12.37524	31.57742
20	16.6987	53.13734204	13.01875	30.74676
54	16.60227	52.86061955	12.964	31.30947
27	16.2511	53.89291984	12.36541	31.25593
82	17.98198	52.77659173	12.03223	30.98149
115	17.8919	52.80838934	12.22811	30.88423
18	17.83404	52.79448665	12.36078	30.83672
103	17.51958	53.30091933	12.28216	30.73292
110	16.16872	54.16448234	12.67058	30.84728
33	17.86258	52.74814763	12.37217	30.91757
69	18.45274	52.80556508	12.10605	30.58287
38	16.93859	52.8086863	12.4512	31.75032
72	16.76732	53.17406799	12.72557	31.34039
35	16.94666	53.18904009	12.48728	31.42965
5	16.60053	52.89982153	12.76585	31.93528
8	19.05901	52.14743699	12.08976	30.92731
10	19.30488	51.62140666	12.41325	31.05501
80	18.42088	52.26580247	12.56237	31.24808
56	19.31443	52.4369589	12.67857	30.2979
48	18.36104	52.70922383	12.77674	30.88804
68	17.12687	53.75205687	12.57437	31.29893
1	18.28627	52.17542211	12.60772	31.77799
83	18.0015	52.93241052	12.57792	31.37189
101	19.14475	52.94006601	12.34178	30.46633
106	17.28314	53.78725225	12.77865	31.08762
92	17.94772	53.7633324	11.95688	31.37389

Rule-set	Tenure	Accommodation Type	Ethnicity	Age
2	18.76227	52.4877037	12.18568	31.67003
90	18.84346	53.50942586	12.16945	30.60523
47	19.18736	52.81908992	12.40614	30.76948
50	18.96659	52.99096056	12.13044	31.10804
29	19.28208	52.54754234	12.50564	30.89144
23	16.75684	54.00869764	12.76855	31.71391
31	18.3515	53.4193322	12.2168	31.29782
71	19.92267	52.24731488	12.56048	30.55886
9	18.52343	52.44019331	12.95844	31.38446
30	19.08471	52.44674232	12.18565	31.60183
19	19.06002	52.11386909	12.92972	31.23362
99	18.40426	52.68755708	13.07966	31.22762
37	19.64318	52.7155606	12.74525	30.3426
16	18.41792	53.12761791	12.47077	31.45923
34	19.54549	52.54885851	12.61101	30.8128
14	19.65266	52.22704667	12.76826	30.92828
73	18.62811	52.78669786	12.07129	32.09853
4	19.31497	52.02896868	12.58069	31.67136
7	19.33999	53.05194112	12.60014	30.7172
121	17.87862	53.90026802	12.92722	31.06263
44	19.96585	52.10418944	12.69875	31.03899
13	19.41163	53.01406907	12.41528	31.02092
25	19.09721	53.11445893	12.76791	30.99634
3	19.53483	52.80187153	13.03499	30.61177
70	19.97332	52.81321933	12.40127	30.91441
45	18.45576	53.48916351	12.42676	31.83144
15	19.49178	52.31611001	12.72	31.85586
93	18.26018	53.93054476	12.42829	31.81793
97	19.08368	55.665819	12.35942	29.45699
22	19.34837	53.31293639	12.7836	31.15043
40	19.48032	52.86265972	12.79087	31.58491
88	16.86566	56.4471338	12.18237	31.53678
98	17.89975	53.73530889	12.48782	33.11773
41	16.81582	56.73140449	12.27329	31.68787
113	17.16491	57.4156603	11.86249	31.19674

Rule-set	Tenure	Accommodation Type	Ethnicity	Age
112	16.98822	57.0722421	12.28448	31.34812
57	16.77133	56.81111488	12.27037	32.00147
108	18.43023	56.41761093	11.56979	31.47084
26	16.49392	57.55080789	11.76531	32.08205
63	16.78764	57.89509027	11.75963	31.51487
79	18.79958	55.9964001	11.73532	31.56714
125	16.88321	57.63345166	12.52007	31.2017
102	18.53548	55.9106515	12.16921	31.63584
126	18.46363	56.08906738	12.11742	31.61207
67	16.45296	57.13563115	12.43113	32.33684
124	18.41427	56.74403729	12.06239	31.22609
60	16.47812	57.38885828	12.46751	32.1615
95	16.87039	57.80222596	12.52329	31.30758
123	18.43951	56.47499918	12.4065	31.19592
117	18.40494	56.80024879	12.10787	31.21228
78	16.78582	56.70641748	12.41454	32.6672
118	18.76543	55.96898455	12.0019	31.85094
105	18.96188	56.00291549	12.40412	31.29247
76	16.96513	57.97641224	12.15437	31.74393
86	17.13431	57.43564564	12.33066	32.09057
119	16.55733	57.87020492	12.09402	32.48628
96	18.33066	56.59975411	11.9251	32.2193
114	18.406	55.93109421	12.53182	32.52083
85	19.73541	56.1296887	11.8363	31.6938
43	19.67376	56.44045977	12.02187	31.32009
51	18.43378	56.46084321	12.24844	32.32279
89	18.48294	56.62883911	12.43241	31.97927
127	18.58478	56.56144496	12.19205	32.25921
64	20.20873	56.31574956	12.33524	30.9168
109	17.06295	57.66910523	12.4967	32.55566
122	20.76704	56.13176025	11.68426	31.26281
107	19.9184	56.66089436	12.18804	31.18211
87	20.57191	55.54934897	12.47219	31.36618
55	19.65012	56.22471672	12.28299	31.84012
53	20.2308	56.24837785	11.7801	31.75217
6	19.40715	56.41863467	12.43067	31.88943

Rule-set	Tenure	Accommodation Type	Ethnicity	Age
65	19.12443	56.10902681	12.54639	32.42489
120	18.9357	56.22942078	12.59394	32.46766
12	19.4769	56.53422609	12.45541	31.81458
49	20.54206	56.12786272	12.21679	31.41213
21	19.41343	56.38197895	12.68973	31.92733
75	20.38574	56.86679072	11.84347	31.5344
17	20.71765	55.92163603	11.84692	32.21148
116	20.89783	56.31815561	12.03326	31.45155
66	20.55683	56.20782545	11.91002	32.029
32	20.44813	56.3812459	11.87717	32.02522
111	19.18541	56.97709546	12.6376	31.96411
24	20.01172	56.40434042	12.31144	32.15546
39	19.59724	56.61158481	12.36015	32.41227
104	20.53447	56.5079196	12.32486	31.76171
36	19.98834	56.69211992	12.32306	32.25033
84	20.72229	56.58163077	12.46487	31.54554
28	20.56469	56.77853309	12.30759	31.6649
100	21.04589	55.66995492	11.84404	32.76216
46	20.79003	56.02647998	12.43572	32.13493
59	20.4443	57.14055797	12.2222	31.77039
62	20.15736	57.39579615	12.24274	31.78222
74	20.21173	57.23307729	12.3798	31.75801
77	20.627	56.4388205	12.20134	32.43562
94	21.12694	57.09510176	12.42469	32.94276

F Mann-Whitney U Test Calculation

	Output Areas	Diversity Index	Rank (Baseline)	Rank (Roads Scenario)
1	00DAFF0047_R	0.49		1
2	00DAFF0047_B	0.52	2	
3	00DAFF0058_R	0.57		3
4	00DAFF0023_R	0.6		4
5	00DAFF0058_B	0.62	5	
6	00DAFF0023_B	0.63	6	
7	00DAFF0062_R	0.73		7.5
8	00DAFF0044_R	0.73		7.5
9	00DAFF0062_B	0.75	9	
10	00DAFM0002_R	0.77		10.5
11	00DAFM0002_B	0.77	10.5	
12	00DAFF0008_R	0.78		12.5
13	00DAFF0007_B	0.78	12.5	
14	00DAFF0044_B	0.79	14.5	
15	00DAFM0015_R	0.79		14.5
16	00DAFF0007_R	0.8		16.5
17	00DAFM0060_B	0.8	16.5	
18	00DAFM0015_B	0.8	16.5	
19	00DAFM0025_R	0.8		16.5
20	00DAFF0008_B	0.8	16.5	
21	00DAFM0025_B	0.8	16.5	
22	00DAFM0060_R	0.81		22.5
23	00DAFM0064_B	0.81	22.5	
24	00DAFF0057_R	0.82		24.5
25	00DAFF0057_B	0.82	24.5	
26	00DAFM0064_R	0.82		24.5
27	00DAFF0012_R	0.82		24.5
28	00DAFM0058_B	0.82	24.5	
29	00DAFF0003_B	0.83	29.5	
30	00DAFF0036_B	0.83	29.5	
31	00DAFF0056_B	0.83	29.5	
32	00DAFF0012_B	0.83	29.5	
33	00DAFF0003_R	0.84		33
34	00DAFF0036_R	0.85		34.5
35	00DAFM0058_R	0.85		34.5

	Output Areas	Diversity Index	Rank (Baseline)	Rank (Roads Scenario)
36	00DAFF0056_R	0.85		34.5
37	00DAFM0019_B	0.85	34.5	
38	00DAFF0029_B	0.85	34.5	
39	00DAFF0029_R	0.86		39
40	00DAFM0019_R	0.87		40
	Total		384	405

Here the notation 'output area_B' represents the baseline results while 'output area_R' represented the scenario 2 results.

G Index of Segregation

Baseline Results (Diversity Indices)

LLSOA Name	LLSOA	Output Area	DI (OA) 2001	DI (OA) 2011	DI (OA) 2021
Seacroft	E01011661	00DAGE0048	0.369476	0.428089	0.40449
Harehills	E01011675	00DAGF0069	0.448278	0.556372	0.594462
Gipton	E01011346	00DAFF0047	0.452016	0.522949	0.539153
Seacroft	E01011658	00DAGE0011	0.485842	0.556372	0.594462
Harehills	E01011679	00DAGF0066	0.501773	0.572609	0.593874
Richmond Hill	E01011626	00DAGB0045	0.511234	0.595369	0.594388
Richmond Hill	E01011619	00DAGB0015	0.516918	0.617319	0.595741
Gipton	E01011346	00DAFF0023	0.540527	0.629465	0.629921
Seacroft	E01011658	00DAGE0012	0.598622	0.638673	0.609389
Gipton new development	E01011431	00DAFM0025	0.798564	0.800628	0.765184

Baseline Results (Index of Segregation)

LLSOA Name	LLSOA	Output Area	IoS (LLSOA) 2001	IoS (LLSOA) 2011	IoS (LLSOA) 2021
Seacroft	E01011661	00DAGE0048	0.165302	0.405438	0.278542
Harehills	E01011675	00DAGF0069	0.085371	0.271583	0.270588
Gipton	E01011346	00DAFF0047	0.147786	0.311489	0.384946
Seacroft	E01011658	00DAGE0011	0.094209	0.4171	0.478283
Harehills	E01011679	00DAGF0066	0.142513	0.244401	0.289797
Richmond Hill	E01011626	00DAGB0045	0.126839	0.188037	0.34569
Richmond Hill	E01011619	00DAGB0015	0.192483	0.559073	0.619359
Gipton	E01011346	00DAFF0023	0.147786	0.311489	0.384946
Seacroft	E01011658	00DAGE0012	0.094209	0.4171	0.478283
Gipton new development	E01011431	00DAFM0025	0.128116	0.397547	0.39002

Scenario 1 Results (Index of Segregation)

LLSOA Name	LLSOA	Output Area	IoS (LLSOA) 2001	IoS (LLSOA) 2011	IoS (LLSOA) 2021
Seacroft	E01011661	00DAGE0048	0.165302	0.366643	0.357078
Harehills	E01011675	00DAGF0069	0.085371	0.294595	0.292242
Gipton	E01011346	00DAFF0047	0.147786	0.338276	0.352847
Seacroft	E01011658	00DAGE0011	0.094209	0.436366	0.485524
Harehills	E01011679	00DAGF0066	0.142513	0.29033	0.310476
Richmond Hill	E01011626	00DAGB0045	0.126839	0.435381	0.589384
Richmond Hill	E01011619	00DAGB0015	0.192483	0.610947	0.667684
Gipton	E01011346	00DAFF0023	0.147786	0.338276	0.352847
Seacroft	E01011658	00DAGE0012	0.094209	0.436366	0.485524
Gipton new development	E01011431	00DAFM0025	0.128116	0.365593	0.361638

Scenario 2 Results (Index of Segregation)

LLSOA Name	LLSOA	Output Area	IoS (LLSOA) 2001	IoS (LLSOA) 2011	IoS (LLSOA) 2021
Seacroft	E01011661	00DAGE0048	0.165302	0.30218	0.316004
Harehills	E01011675	00DAGF0069	0.085371	0.119925	0.18779
Gipton	E01011346	00DAFF0047	0.147786	0.307197	0.435531
Seacroft	E01011658	00DAGE0011	0.094209	0.424311	0.423905
Harehills	E01011679	00DAGF0066	0.142513	0.237178	0.259148
Richmond Hill	E01011626	00DAGB0045	0.126839	0.525041	0.576217
Richmond Hill	E01011619	00DAGB0015	0.192483	0.61719	0.66486
Gipton	E01011346	00DAFF0023	0.147786	0.307197	0.435531
Seacroft	E01011658	00DAGE0012	0.094209	0.424311	0.423905
Gipton new development	E01011431	00DAFM0025	0.128116	0.393369	0.481648

Scenario 3 Results (Index of Segregation)

LLSOA Name	LLSOA	Output Area	IoS (LLSOA) 2001	IoS (LLSOA) 2011	IoS (LLSOA) 2021
Seacroft	E01011661	00DAGE0048	0.165302	0.432376	0.348185
Harehills	E01011675	00DAGF0069	0.085371	0.140027	0.182336
Gipton	E01011346	00DAFF0047	0.147786	0.337207	0.361579
Seacroft	E01011658	00DAGE0011	0.094209	0.445678	0.443737
Harehills	E01011679	00DAGF0066	0.142513	0.231599	0.258418
Richmond Hill	E01011626	00DAGB0045	0.126839	0.479455	0.553983
Richmond Hill	E01011619	00DAGB0015	0.192483	0.595282	0.65024
Gipton	E01011346	00DAFF0023	0.147786	0.337207	0.361579
Seacroft	E01011658	00DAGE0012	0.094209	0.445678	0.443737
Gipton new development	E01011431	00DAFM0025	0.128116	0.40102	0.385361

H Multi-Agent-Based Simulation XI Paper

Agent-based Simulation Modelling of Housing Choice and Urban Regeneration Policy

René Jordan, Mark Birkin, Andrew Evans

Centre for Spatial Analysis and Policy

School of Geography, University of Leeds

Leeds, LS2 9JT, UK

Corresponding author r.j.jordan@leeds.ac.uk

Abstract. Phenomena in the housing market can be recreated and analysed using the technique of agent-based modelling. Housing policies, such as urban regeneration, seek to address problems of deprivation in segregated communities by introducing the concept of mixed communities, that is, communities mixed by housing tenure and housing type. In this paper, a framework for the creation of a model of housing choice and regeneration policy is presented.

Keywords: agent-based modelling, housing choice, urban regeneration policy

Introduction

Housing Policy is one of the instruments used by government to manage the housing sector and includes, as a part of its remit, attempts to improve the dwelling conditions of those unable to provide suitable homes for themselves. Deprivation and the state of the poor have played instrumental roles in the direction of these policies. One such policy has been presented under the umbrella of Urban Regeneration. As defined by Bramley *et al.* [1], regeneration is the process of recovering and renewing lost vitality to the physical and social landscape. Hull [2] argues, however, that despite the physical changes in the urban mosaic of most regenerated cities, Urban Regeneration Policy does not effect a narrowing of the gap between the disadvantaged and those of higher social standing. On the contrary, Hull [2] calls the policy a failure.

These contrasting viewpoints raise many questions. Is the government's new goal of equipping the less advantaged with the tools to seek market provisions likely to yield successful results? What are the likely results of the recent housing-led regeneration policies and will these results fall in line with the goals envisioned by government? In this paper, we suggest that agent-based modelling (ABM) is a technique which can illuminate the problems associated with Urban Regeneration Policy.

Housing market models in the realm of social simulation are discussed forming the precursor for the introduction of a new housing market model. An original modelling framework is presented which refines conventional notions of preference to include a broad range of socio-demographic, economic and geographical variables. The importance of model testing and validation in specific local contexts will be emphasized, and an empirical application for the area of East and South-East Leeds (EASEL), England, will be developed.

Urban Regeneration Policy and EASEL

With a population of over 700,000 residents, Leeds is one of the largest metropolitan districts in England. The city is characterised by a booming financial sector and a large student population. Despite this view of the city, it contains some of the most deprived communities in the United Kingdom [3]. At least 46,000 Leeds residents live in areas rated amongst the 3% most deprived in England [4]. Most of these residents live within the EASEL area.

The EASEL area is resident to more than 36,000 households. According to the EASEL Aspiration Needs and Housing Study 2007 [5], 85% of the Super Output Areas (SOAs) in EASEL fall within the top 10% most deprived in England while 91% fall within the top 20% in England. (A SOA is a neighbourhood with approximately 300 households.) Issues of deprivation and social disadvantage, high unemployment, and above average rates of crime plague these communities.

Of primary interest to this research is the role of housing in the regeneration scheme. The central policy objective is the creation of sustainable communities, a term strongly linked to mixed communities. Leeds City Council [4] believes that the success of this goal hinges on the creation of a stable housing market. The council intends to introduce a greater mix of housing tenures in council owned areas by introducing private housing. This, it argues, reduces movement turnover in communities thus providing a gateway for creating sustainable communities. In order to facilitate this, an estimated 7,800 new homes are to be built to create these new mixed communities – mixed by tenure. Note that, mixed communities are communities diversified by socio economic status and housing tenure. In the UK context, housing tenure can be largely divided into two categories; social housing – houses owned and or administered by the Local Council or Private housing.

Proponents for this form of tenure diversification argue that mixed communities can contribute to a smaller concentration of unemployed people by attracting economically active households to previously deprived neighbourhoods [6]. Others claim, however, that though this can thin out the problem of deprivation, it still does not solve the problem of social disadvantage [7], [8].

Whether the policy of creating mixed communities will yield the required results is questionable - there are not sufficient results on which to base an informed judgement. What is known, however, illustrates that the theory overshadows the practicality of the results. Through computational modelling, the validity of these hypotheses can be tested.

Social Simulations in the Housing Market

Approaches to housing market modelling are not new to the field of computer simulation. The dynamics of this market are intricately woven into the complex system of the world in which we live. Its volatility can be seen as house prices fluctuate due to activity in the financial market, affecting terms of lending, interest rates and general attitudes towards risk, among others. Merging these factors with discriminatory individual behaviour creates an environment ripe with modelling opportunities.

Existing research on the dynamics of the housing market is extensive [9], [10], [11], [12], [13], [14], [15]. Issues of residential preferences, ethnic segregation within communities, residential mobility, housing choices and the impact of government policy, continue to generate interest. Through computer models and simulation, the intricacies of this dynamic market can be explored.

The work of Thomas Schelling is noted to be one of the first agent-based models of its kind to replicate discriminatory individual level behaviour in the form of a model [16]. Schelling [17], [12] examined the role of preferences in an artificially created community and illustrates how individual behaviour can create significant collective results not directly intended by the individual [17]. Schelling proved that even with slight preferences, total segregation can be effected if these preferences are exercised.

Schelling's work, though simple, forms the basis for much research on individual choice, segregation and integration. Work by Pancs and Vriend [18] examined the role of preferences in relation to integration policy. They concluded that even when individuals preferred integrated neighbourhoods, the impact of preferences led to segregated communities. Furthermore, when public policies were enacted to heighten tolerance levels, individuals still gravitated towards others like themselves. In a similar way, Zhang [19], [20], in his mathematical model, concludes that even in areas where pure integration is preferred, segregation is likely.

Aguilera and Ugalde [21] attached house prices to each space on a lattice grid. Individuals were rated by socioeconomic status and income and moved to match their status with the price of their house. In this case, segregation was observed. Yin [22] increased the dynamics in his model by devising a social simulation to examine the issue of race, social class and residential segregation. He illustrates that factors such as race and economic constraints, when exercised as a part of the housing choice process, can cause segregation of varying degrees at the aggregate level. However, Yin illustrated that when housing policies were implemented this segregation could be reduced once racial sensitivity was low. Therefore, integration seems likely if people are educated to favour it and/or housing policies are implemented to create integrated societies. However, if communities are left to form naturally with limited interference where policy is concerned, segregation is likely to occur.

Models like these tackle various aspects of the problem of segregation and integration as they relate to activity in the housing market. The model outlined in this paper encapsulates the design noted in previous, similar social simulations while extending the design further to mirror conditions and trends in the EASEL area more closely. Such a real world application is aimed at refining Schelling's notion of preferences to include not only ethnicity but also preferences pertaining to the family life cycle; tenure type; accommodation type; distance to city; accessibility of transport routes; distance to schools; cost of housing and knowledge of the new neighbourhood. These preferences interact with policy directives and environmental conditions such as changes in interest rates, in/out migration and the presence of new facilities such as schools. Not only is there no record of this being done but applying such a model to an existing project such as EASEL provides the opportunity to test model outcomes against actual outcomes as time progresses.

Research such as this challenges our understanding of causal relationships in the housing market and more specifically in the EASEL area. At the aggregate level, policy makers are able to gauge how population profiles change over time, raising a need for more services such as schools and healthcare facilities. In a similar way, such research could point out where services are not used sufficiently and lead to reassessment of resource planning. This is important when asset management is considered, especially amidst the reality of difficult economic times. Also, having never been implemented in the EASEL area, regeneration policies do not have a proven track record. A model such as this can provide a platform for scenarios to be created and tested in an effort to speculate on their performance.

Though other modelling techniques are well established in the housing market domain, such as microsimulation [23] and spatial interaction modeling [24], agent-based modeling allows for the manipulation of individual level behavior at an atomic level. Agent-based systems appear to recreate events in ways more similar to activities in the real world. They are dynamic in nature as agent states continue to change due to their interaction with other agents and interaction with the environment in which they exist. Collective resultant behaviour possibly characterised by emergence may provide further useful insights beyond conventional results [25]. Building on this premise, the framework for the EASEL Housing Simulation model is discussed.

The Model Defined

As a replica of activity in the EASEL area, the EASEL Housing Simulation uses households and houses to represent individual agents. For the purpose of this project, a household is used to represent a collection of residents living together. Details of the household representative person were derived from the Household Sample of Anonymised Records therefore one record is used to represent the entire household. Such a record is deemed sufficient in representing this unit as it contains details on the number of residents in the household, including the number of children.

Fundamentally, each household resides in a house for an undetermined time period until some push factor, influences the decision to move. These push factors may range from changes in household size, to changes in disposable income or forced moves of council tenants as initiated by the City Council. In general, household agents are inputted into the simulation and initially assigned to houses. As the model advances from one time step to the other, environmental variables are updated to simulate changing economic and social conditions in the market. While this happens, households wishing to move are identified and attempts to find a suitable new dwelling is made.

The underlying framework of this model is presented in the sections to follow. We examine the key stages in this process, beginning with the derivation of the input data, and the assignation of households to housing. Time stepping in the model and the determination of movers is explained. Then the location decisions of households are examined along with the background modelling of environmental variables.

Derivation of the Input Data

In demographic terms, the starting point for the simulation is a complete representation of households in the EASEL area. Starting with a large sample from the UK population, households are selected to match the characteristics of EASEL (for example, high levels of council-owned housing, significant deprivation) using a reweighting process which is well-known in the spatial microsimulation literature [26]. Household data is generated from the Household Sample of Anonymised Records (SAR) for England and Wales (www.ccsr.ac.uk), and output area data is accessed from the Census Area Statistics. This method provides a complete representation for individual households of attributes collected in the UK Census, including ethnicity, age, family composition, health status, accommodation type and housing tenure. This range of attributes provides the basis for implementing a rich set of rules for household movement and destination choice. Shapefiles representing houses and roads are derived from data provided through the Ordnance Survey, while Output Area boundaries are downloaded through Edina UK Borders (www.edina.ac.uk).

Assignation of Households to Houses

Attached to each Output Area (OA) is a set of attributes describing it. Crime rates, house quality, access to healthcare etc are encapsulated in the Index of Multiple Deprivation variable. The Index of Multiple Deprivation rates the level of deprivation for each Output Area. Better areas are characterised by higher IMD figures. When households are initially read into the model, each household is assigned to a house using the OA field in the household record. This matching of OA and households ensures that households are placed in areas which match their socio economic status. Note that this process is necessary as data produced during the microsimulation only contains OA references and not exact house codes. This is important in order to ensure that actual individuals cannot be identified in the Census data.

Time Stepping

The technique of time stepping is used in the simulation project to recreate an environment where events are measured in actual time. In this way, the simulation can mimic time driven events in the real world. We choose to increment the time step counter on a yearly basis as immigration rates can be monitored at this level.

Determination of Household Movers

The probability that a household wishes to move in a specific time interval is derived through an analysis of the Household SAR, which includes the variable 'moved in the last year' alongside other social, demographic and household characteristics. In order to determine different movement probabilities for different socio-demographic groups, we built a decision-tree using the SPSS AnswerTree extension (www.spss.org). The Household SAR contains several categorical variables, AnswerTree allows for the manipulation of such through the use of chi-square analysis.

The decision tree is shown schematically in Figure 1. At each level in the tree, a household attribute is identified which differentiates by movership. For example, at the first level in the tree, Branch 1 represents household heads aged 25-44 (high movement), branch 2 is ages 45-64 (moderate movement), and branch 4 is ages 65+ (low movement). Branch 3 represents young adults (under the age of 25), with very high levels of movement. At the next level of the tree, each branch is further sub-divided by the next differentiating attribute. In the case of young adults, there is a further three-way split which is based on housing tenure (e.g. private renters have the highest rates of movement). The process continues for as long as significant factors can be identified to differentiate migration probabilities between households.

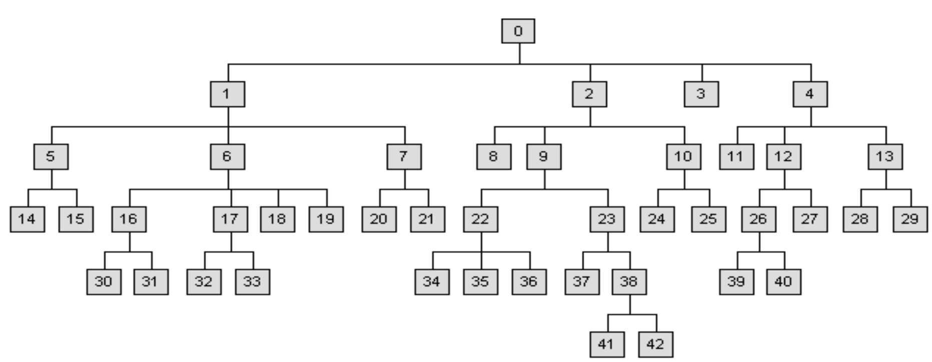


Fig. 1. Decision Tree

At the bottom of the tree this leads to clusters of households with discrete characteristics and an associated rate of movement. When the simulation is implemented, the characteristics of a household are parsed in order to allocate the household to an appropriate cluster. Thus a household with a head aged between 25 and 44, renting privately with less than 6 rooms would be allocated to cluster 14, with a migration probability of 0.53. A random number is generated in the simulation, and if that number is lower than 0.53 then this household will be directed to the movement pool in the simulation.

A full list of 22 clusters from the decision tree is shown in Table 1. It can be seen that age and household composition; housing tenure, size, and accommodation type; and the occupation of householders are all important drivers of the movement process. Observe that notable attributes such as ethnicity (and others) were considered in the decision tree but not found to be significant. We conclude that any variations in movement between ethnic groups are proxied by other variables such as household size and tenure, but also note that ethnicity can still be important in the choice of destination, which is a separate process (see below). Note that, a variable is classified as significant when there are at least 1000 cases available for examination in the entire dataset. Branches are therefore terminated when this condition is not met.

Table 1. Tabular list of decision tree clusters

Cluster code	Node	Level 1	Level 2	Level 3	Level 4
		Age of Head	Tenure	Housing reqd	
1a1	14	▪Adults 26-45	▪Private rented	▪<6 rooms	xxxxxxxx
1a2	15			▪>5 rooms	xxxxxxxx
				Household comp.	
1b1	18			▪Single	xxxxxxxx
1b2	19		▪Owner-occ	▪Couple	xxxxxxxx
					Crowding
1b3i	30			▪Family	▪Low density
1b3ii	31				▪Highdensity
					Socio-economic status
1b4i	32			▪Lone parents	▪Low status
1b4ii	33				▪High status
				Household comp.	
1c1	20		▪Council rented	▪Lone parent	xxxxxxxx
1c2	21			▪Single/couple/family	xxxxxxxx
			Tenure		
2a	8	▪Adults 46-65	▪Private rented	xxxxxxxx	xxxxxxxx
				Housing reqd.	House size
2b1i	34		▪Owner occupied	▪Rooms required 0-10	▪Rooms occupied 1-5
2b1ii	35				▪Rooms occupied 6-10
2b1iii	36				▪Rooms occupied 11+
					▪Household comp.
2b2i	41			▪Rooms required 11+	▪Family
2b2ii	42				▪Other hh comp
3	3	▪Adults 17-25	xxxxxxxx	xxxxxxxx	xxxxxxxx
			Tenure		
4a	11	▪Adults 65+	▪Flat	xxxxxxxx	xxxxxxxx
				Household comp.	
4b1	39		▪Terrace	▪Married couple/family	xxxxxxxx
4b2	40			▪Other hh comp	xxxxxxxx
4c1	28		▪Detached house	▪Married couple/family	xxxxxxxx
4c2	29			▪Other hh comp	xxxxxxxx

Selection of Destinations

When it is determined that a household will move, a new home must be found. Research examines the type of behaviours contributing to residential segregation as a consequence of selecting a location at which to live [10], [13], [14]. Segregation is often attributed to issues such as culture, religion, language, ethnicity groups, economic advantage and school searches [14]. Forces for segregation and dispersal are complex, dynamic and contextual in that they are experienced in different ways in different places by different types of households [14].

The choice of where to move to is a combination of dwelling alternatives. One of the most important factors in this decision is the financial budget of the household [27], [28]. This is integral to the decision to be made as across the housing market, houses are grouped according to price. Even when limited by a budget, neighbourhood characteristics are important factors as they influence where household moves. The physical conditions of the neighbourhood, amenities such as shops, school quality, security and transport connections are characteristics which can determine whether a household chooses to live in an area or not. Importantly, the significance of each of these variables hinges not only on the preferences of the household but also on the household's ability to afford the new dwelling as noted earlier.

In a similar way, dwelling characteristics are important. Dwelling characteristics such as dwelling size, type, age and quality must be included because at various stages of the family life cycle there are different dwelling requirements [29]. The choice to live in a house as opposed to a flat could be the result of a household with children desiring a garden for children to play. Similarly, the number of rooms may be linked to the size of the family. Behaviours such as these can be further extended as represented in the literature. We opt to implement nine main rules to represent the process adopted by households of choosing a new dwelling. The rules have been derived from the existing literature as well as information given during informal talks with personnel from the Leeds City Council and will be implemented individually in the first instance in order to reduce the complexity of interpreting the results. Following this, rules will be weighted according to the scenarios chosen, for example, a poor household, facing unemployment is likely to sacrifice living in near to a community where 'better' schools can be found because such a household may not be in a position to afford the homes found in such an area. Thus in satisfying the conditions for Rule 8, Rule 9 becomes less important.

The rules are presented as follows:

1. Households first tend to look for a new house within known areas [30].
2. Households will move to houses where the size of the house is adequate [30].
3. Households will move to houses where the tenure type of the house is desired [29].
4. Households will move to areas where the ethnic makeup is tolerable [10], [13], [14], [15], [31].
5. Households will move to areas where transport routes are accessible [29], [32].
6. Households containing school-aged children will try to move to areas where better schools are accessible [35], [36], [37].
7. Households will move to houses where the neighbourhood quality is better. [30]

What are the Environmental Variables?

As in the real world, the dynamics of the housing market continually change; interest rates, monetary policies, mortality and fertility rates. Each of these variables has an effect on residential mobility. For example, the birth of a child may lead to a young couple purchasing a new home to accommodate the growing family, whereas the death of an elderly person may result in a spouse opting to buy a smaller home. Changes in interest rates affect house prices in the form of increased or decreased mortgage rates and or rental rates. Such changes may cause the household to find a cheaper home or it may encourage the household to move opportunistically to a better home. Each of these factors can encourage or discourage a household choice of a new home.

Results

In order to generate initial results, five experiments were created and executed. The experiments are presented below with a brief discussion and a view of the way forward presented in Sections 5.1 and 5.2 respectively.

Brief Review and Discussion of Results

Using a sample population of 559 households, 606 houses distributed across 6 Output Area zones, various combinations of assumptions were applied in the model and the results observed. Households are assigned to houses using the process as detailed in Section 4.2.

Table 2. Description of Experiments

Exp#	Description	Assumptions	# Moves	# Iterations	Equilibrium	Clustering
1	Pure Schelling	Random movers Ethnic push	9	45	Yes	Strong
2	Ethnic 1	Random movers Ethnic push Ethnic pull (Rule 6; Section 4.5)	10	70	Yes	Strong
3	Ethnic 2	Mover model Ethnic pull (Rule 6; Section 4.5)	15	900	Limited	Weak
4	Ethnic 3	Mover model Ethnic push Ethnic pull (Rule 6; Section 4.5)	7	450	Yes	Strong
5	Housing	Mover model Ethnic push Housing pull (Rule 3; Section 4.5)	1	20	Very Limited	Very Weak

Exp1 is our starting point and represents the original Schelling model. Here an ethnic push is assumed; individuals are motivated to move based on a dislike for the current ethnic mix in their community and opt to move to any other vacant home. Clustering and equilibrium are realised very quickly with the average number of moves per household recorded at 9. Note that equilibrium refers to a stable state where a negligible number of households are observed to be moving. Every iteration of the model represents a time step of unspecified duration in which movement decisions are evaluated.

Exp2 is an augmentation of Exp1. Both an ethnic push and ethnic pull are assumed here; households leave neighbourhoods where the ethnic mixed is not tolerable and find homes in areas where the ethnic mix is tolerable (Section 4.5; Rule 6). Again segregation is realised though after almost twice the number of iterations as previous.

Exp3, 4 and 5 adopt the mover model as discussed in Section 4.4. In Exp3, the combination of the mover model as the push factor and ethnic tolerance as the pull factor resulted in limited clustering. Adding ethnicity to the push factors result in strong clustering after ~300 iterations as observed in Exp4. Exp5 extended from Exp4 by searching for houses with the required amount of rooms (Rule 3, Section 4.5). As Table 2 notes, equilibrium is limited with only few clusters observed.

These observations highlight the contributions of the mover model in limiting the number of times households move but also point out the importance of coupling this model with more subjective preferences such as ethnicity. Though the ethnicity attribute was not highlighted in the overall model, these experiments show that it is important that provisions be made for its inclusion. Therefore, the results in Exp4 appear to be closer to reality.

Adding more pull factors, as in Exp5, appear to grind the model to a halt very quickly while a negligible number of clusters appear. This is likely to be the case as a result of the limited availability of the required house size, tenure type etc. This is similar to reality as oftentimes a household may compromise on their housing requirements because of limitations in the available housing stock.

Extending these experiments even further, we can compare the results to possible behaviours in reality. Using the mover model to govern the selection of households wanting to move and running the rules in isolation of each other, each of the 7 rules can be implemented to seek results. As expected, the ethnicity rule results in clustering of households according to the ethnic group, a recreation of Schelling's model but in an actual geography. The known area rule results in households moving short distances. The IMD rule sees deprived areas being vacated. With regards to the transport routes and schools rule, households without cars and households with dependent children are seen gravitating closer to these respective resources. While in the case of the rooms and socioeconomic status rules, households strongly limit their selection of a new house to their preference for a specific number of rooms and housing tenure respectively.

We can alter these experiments by coupling the mover model is coupled with the full complement of rules. This means that each of the 7 rules must be satisfied in order for a household to move to a new home. Such a simulation results in limited movement because of the rigidity of the conditions. Even in reality households are known to compromise on some preferences once the decision to move has be made. It therefore means that for each household type a different combination of rules is likely to be more applicable. For example, ethnic minority groups in the EASEL area are known to cluster together strongly and often do no display wealth by way of housing tenure [15]. For such a group the ethnicity rule would be very important while the socioeconomic rule may be ignored. In a similar way, for young students, though small dwellings may be required because of the availability of house shares, they may choose to live in larger houses than necessary. Here the rooms rule is compromised.

Through the process of calibration, rule sets can be created which mirror the behaviours of each type of household. de Smith *et al.* [38] defines calibration as the process of ensuring that the model parameters match the parameters used in the real world to effect real world results. It is the process of refining the behavior of the model to ensure that the model replicates behaviours in the real world. The model is executed with various combinations of rules, the results generated are compared to known results for the EASEL area to test the level of accuracy. Once this process is complete, a collection of rule sets for each household type should be identifiable leading to the generation of final model results.

Data rich sources such as the National Shoppers' Survey and the Pupil Level Annual School's Census (PLASC) are used in this calibration process to access the extent to which the model is able to recreate reality. The National Shoppers' Survey contains demographic details and preference data of householders while the PLASC contains demographic details on school children. Counts of household by ethnicity, age etc. are compared and the goodness of fit indicator, the Index of Heterogeniety, are used to assess the level of heterogeniety in each Output Area. Blau [39] uses the following notation to describe the index:

$$H = 1 - \sum P_i^2$$

Where P is the proportion of a particular group in each area i. The Index of Heterogeneity returns a value between 0 and 1 and is defined at the Output Area level of our model. A value closer to 0 denotes a high level of segregation while a value closer to 1 denotes a high level of diversity. Values generated from our model can be compared to values generated from the real data.

Where to Next?

With this framework in place, the model can then be used to run scenarios; new houses can be added to the EASEL landscape, schools may be moved and or houses demolished to examine the likely effects of Regeneration Policy. Other data sources such as the British Household Panel Survey (BHPS) may also be used to assist in updating the environmental variables of the model. Such a dataset is similar to the Household SAR in that it is devoid of spatial references but information rich. Finally, the model in its present state may be described as closed; households neither move beyond the boundaries of EASEL nor enter from outside this area. It may be argued that this is acceptable, that is, though households move, the same type of households exist in therefore moving them around in the same output space is likely to generate the same or similar results as an open system. However, scaling up the model to represent Leeds is another option, though the computational requirements for such a venture may outstrip the capacity of the available systems.

Conclusion

The housing market is stratified, so without policies which support mixed communities, households will cluster according to socio economic status and ethnicity. In this paper, we have established a policy framework for housing market behavior with urban regeneration. A rich basis for the creation of agents and their movement patterns has been introduced. We set out rules for location decisions based on a diverse set of characteristics and migration behaviours. From our early experiments to explore the effect of agent rules, further results and policy simulations are now awaited which can begin to support the policy process and provide real insights into the establishment and maintenance of socially mixed communities.

Acknowledgements

Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

References

1. Bramley, G., Pawson, H., Munro, M.: Key Issues in Housing: Policies and Markets in 21st Century Britain. Palgrave Macmillan, Britain (2004).
2. Hull, A.: Neighbourhood Renewal: A toolkit for regeneration. GeoJournal. 51(4), 301-310 (2000).
3. Unsworth, R., Stillwell, J.: Twenty-First Century Leeds: Geographies of a Regional City Rev Ed., PLACE Research Centre, York (2004).
4. Leeds City Council. East and South East Leeds Area Action Plan: Leeds Local Development Framework (2007).
5. Leeds City Council. EASEL Aspiration Needs and Housing Study 2007 (2007).
6. Beekman, T., Lyons, F., Scott, J.: Improving the understanding of the influence of owner occupiers in mixed tenure neighbourhoods. ODS Ltd for Scottish Homes, Edinburgh (2001).
7. Kleinhans, R.: Social implications of housing diversification in urban renewal: A review of recent literature. Journal of Housing and the Built Environment. 19(4), 367-390 (2004).
8. *Kintrea*, K.J., *Atkinson*, R.: Owner-occupation, social mix and neighbourhood impacts. Policy & Politics. 28(1), 93-108 (2000).
9. Gilbert, A.: A home is forever? Residential mobility and homeownership in self-help settlements. Environment and Planning A. 31(6), 1073-1091 (1991).
10. Johnston, R.J., Poulsen, M.F., Forrest, J.: Are There Ethnic Enclaves/Ghettos in English Cities?. Urban Studies. 39: 591–618 (2002).
11. Kim, J., Pagliara, F., Preston, J.: The Intention to Move and Residential Location Choice Behaviour. Urban Studies, 42(9), 1621-1636 (2005).
12. Schelling, T.: Dynamic models of segregation. Journal of Mathematical Sociology. 1(2), 143-186 (1971).
13. Peach, C.: London and New York: contrasts in British and American models of segregation. International Journal of Population Geography. 5, 319–351 (1999).
14. Phillips, D.: Black Minority Ethnic Concentration, Segregation and Dispersal in Britain. Urban Studies. 35(10), 1681-1702 (1998).
15. Phillips, D.: Ethnic and Racial Segregation: a critical perspective. Geography Compass. 1(5), 1138-1159 (2007).
16. Macy, M.W., Willer, R.: From factors to actors: computational sociology and agent-based modeling. Annual Review of Sociology. 28 (2002).
17. Schelling, T.: Models of Segregation. The American Economic Review. 59(2), 488 – 493 (1969).
18. *Pancs*, R., *Vriend*, N.: Schelling's Spatial Proximity Model of Segregation Revisited. Journal of Public Economics. 91, 1-24 (2007).
19. Zhang, J.: A Dynamic Model of Residential Segregation. The Journal of Mathematical Sociology. 28(3), 147-170 (2004).
20. Zhang, J.: Residential segregation in an all-integrationist world. Journal of Economic Behavior & Organization. 54(2004), 533-550 (2004).
21. Aguilera, A., Ugalde, E.: A Spatially Extended Model for Residential Segregation. Discrete Dynamics in Nature and Society. Article ID 48589, 20 pages (2007).
22. Yin, L.: The Dynamics of Residential Segregation in Buffalo: An Agent-based Simulation. Urban Studies. 46(13), 2749-2770 (2009).
23. Miller, E., Hunt, H., Abraham, J., Salvini, P.: Microsimulating urban systems. Computers, Environment and Urban Systems. 28, 9-44 (2004).
24. Leishman, C., Bramley, G.: A local housing market model with spatial interaction (2005).
25. Epstein, J.M.: Agent-based computational models and generative social science. Complexity. 4(5), 41-60 (1999).

26. Williamson, P., Birkin, M. Rees, P.H.: The estimation of population microdata using data from small area statistics and samples of anonymised records. Environment and Planning A. 30, 785-816 (1998).
27. Tu, Y., Goldfinch, J.: A Two-stage Housing Choice Forecasting Model. Urban Studies. 33(3), 517-537 (1996).
28. Boehm, T.: A Hierarchical Model of Housing Choice. Urban Studies. 19(1), 17-31 (1982).
29. Dieleman, F.: Modelling residential mobility; a review of recent trends in research. Journal of Housing and the Built Environment. 16(3), 249-265 (2001).
30. Cho, Y., Who moves and where? A comparison of housing association tenants in London and northern regions. Housing Corporation London and Sector Study 40. Housing Corporation: London (Nov 2004b).
31. van *Kempen*, R., *Özüekren*, A.: Ethnic segregation in cities: new forms and explanations in a dynamic world. Urban Studies. 35, 1631-1656 (1998).
32. Forrest, R.: Spatial Mobility, tenure mobility, and emerging social divisions in the UK housing market. Environment and Planning A, 19, 1611-1630 (1987).
33. Cho, Y., Lyall Grant, F., Whitehead, C., Affordable housing in London: who expects to move and where? Housing Corporation Sector Studies: Sector Study 39. Housing Corporation: London (Nov 2004a).
34. Böheim, R., Taylor, M.P.: Residential Mobility, Housing Tenure and the Labour Market in Britain, ILR working papers 035, Institute for Labour Research (1999).
35. Gibbons, S., Machin, S.: Valuing English Primary Schools', Journal of Urban Economics. 53, 197-219 (2003).
36. Black, S.: Do Better Schools Matter? Parental Valuation of Elementary Education. Quarterly Journal of Economics, 578-599 (1999).
37. Strand, S.: Pupil Mobility, Attainment and Progress in Key Stage 1: A Study in Cautious Interpretation. British Education Research Journal. 28, 63-78 (2002).
38. de Smith, M., Goodchild, M., and Longley, P.: Geospatial Analysis. Mathador: United Kingdom (2009).
39. Blau, P.: Inequality and Heterogeneity: A Primitive Theory of Social Structure. Free Press: New York (1977).

I EASEL LLSOAs Defined

Note that the ID field in this table is related to the Ids noted in **Figure 5.1**.

ID	LSOA	Name	"Name" of area	Wards
1	E01011338	Harehills	Haselwoods / Rigtons	Burmontofts
2	E01011339	Gipton	Oaktrees / Beech Mount Oakwood Lane	Burmontofts
3	E01011340	Gipton	Coldcoat Avenue / Kitcheners / Bullers / St Albans	Burmontofts
4	E01011341	Seacroft	Veritys / Dunhills	Burmontofts
5	E01011342	Gipton	Brander Road / South Farms / Coldcotes	Burmontofts
6	E01011343	Seacroft	The Oval	Burmontofts
7	E01011344	Harehills	Bellbrooke Avenue / Kimberley Road / Comptons	Burmontofts
8	E01011345	Seacroft	Crossgates	Burmontofts
9	E01011346	Gipton	Wykebeck Valley Road / Branders / Gipton Approach	Burmontofts
10	E01011347	Harehills	Cliftons / Nowells	Burmontofts
11	E01011348	Harehills	Torres	Burmontofts
12	E01011349	Harehills	Glenthorpes / Gargrave Place / Brignall Garth	Burmontofts
13	E01011421	Gipton	Hollin Parks	Harehills
14	E01011422	Harehills	Markham Avenue / Brookfield Avenue / Roundhay	Harehills
15	E01011423	Gipton	Lawrences / Ambertons / Fearnvilles	Harehills
16	E01011424	Gipton	Hetton Road / Amberton Road / St Wilfrids Crescent	Harehills
17	E01011425	Gipton	Grange Parks	Harehills
18	E01011426	Harehills	Gathorne Terrace / Hares Avenue / Pasture Road	Harehills
19	E01011427	Gipton	Easterly Grove / St Wilfrids	Harehills
20	E01011428	Harehills	Harehills Road / Conway Drive / Luxors	Harehills
21	E01011429	Harehills	Spencer Place / Blankside Street / Shepherds Lane	Harehills
22	E01011430	Harehills	Darfield Road / Sandhursts / Dorsets	Harehills
23	E01011431	Gipton	Foundrys / Thorn Drive / North Farm Road	Harehills
24	E01011432	Harehills	Chatsworth Road / Berkeleys / Strathmore Terrace	Harehills
25	E01011433	Harehills	Comptons / Ashtons / Cowpers	Harehills
26	E01011434	Harehills	Ashtons / Conways	Harehills
27	E01011615	Halton Moor and Osmondthorpe Area	Dawlishes / Skeltons	Richmond Hill
28	E01011616	Halton Moor and Osmondthorpe Area	Carden Avenue / Oak Road / Partage Crescent	Richmond Hill
29	E01011617	Halton Moor and Osmondthorpe Area	Rookwoods	Richmond Hill

ID	LSOA	Name	"Name" of area	Wards
31	E01011619	Richmond Hill	East St / Upper Accommodation Rd / Lavendar Street	Richmond Hill
32	E01011620	Halton Moor and Osmondthorpe Area	Halton Moor / Ullswater Crescent / Rathmell Road	Richmond Hill
33	E01011621	Halton Moor and Osmondthorpe Area	Ings Road / Nevilles / Osmondthorpe Lane	Richmond Hill
34	E01011622	Halton Moor and Osmondthorpe Area	Neville Road / Wykebecks	Richmond Hill
35	E01011623	Richmond Hill	East Park Drive / Glensdales / Raincliffes	Richmond Hill
36	E01011624	Halton Moor and Osmondthorpe Area	Halton Moor / Kendal Drive / Cartmell Drive	Richmond Hill
37	E01011625	Richmond Hill	St Hildas / Copperfields / Gartons	Richmond Hill
38	E01011626	Richmond Hill	Corss Green Lane / Easy Road / Dial Street / Dent Street	Richmond Hill
39	E01011656	Seacroft	Boggart Hill Drive / Barncroft Road / Ramshead Drive	Seacroft
40	E01011657	Seacroft	Ramsheads / Limewoods / Monkswoods	Seacroft
41	E01011658	Seacroft	Boggart Hill	Seacroft
42	E01011659	Seacroft	Kentmere Avenue / North Parkway / Easdales Crescent	Seacroft
43	E01011660	Seacroft	Kentmere Approach / Rosgill Drive / Brooklands Lane	Seacroft
44	E01011661	Seacroft	Eastdeans / Seacroft Crescent / Hansbys	Seacroft
45	E01011662	Seacroft	Foundry Mill Terrace / Brooklands	Seacroft
46	E01011663	Seacroft	Tarnside Drive / Foundry Mill Street / South Parkway	Seacroft
47	E01011664	Seacroft	Redmires / South Parkway / Kentmerre Avenue	Seacroft
48	E01011665	Seacroft	Inglewood Drive / Crossgates Avenue / Stocks'	Seacroft
49	E01011666	Seacroft	Hawkhills / Bryan Crescent / Sandway	Seacroft
50	E01011667	Seacroft	Foundry Mill Drive / Hawkshead Crescent / Alston Lane	Seacroft
51	E01011671	Harehills	Cambridge Road / Servias / Meanwood Road*	University
52	E01011673	Harehills	Bayswaters / Gledows	University
53	E01011675	Harehills	Lincoln Green	University
54	E01011677	Harehills	Shakespeares / Bexleys / Bayswaters	University
55	E01011679	Harehills	Little London / Lovell Park*	University