Geography matters

Available in alternative formats

This publication can be provided in alternative formats, such as large print, Braille, audiotape and on disk. Please contact: Communications Department, Joseph Rowntree Foundation, The Homestead, 40 Water End, York YO30 6WP. Tel: 01904 615905. Email: info@jrf.org.uk

Geography matters

Simulating the local impacts of national social policies

Dimitris Ballas, David Rossiter, Bethan Thomas, Graham Clarke and Danny Dorling

JOSEPH ROWNTREE
FOUNDATION

The **Joseph Rowntree Foundation** has supported this project as part of its programme of research and innovative development projects, which it hopes will be of value to policy makers, practitioners and service users. The facts presented and views expressed in this report are, however, those of the authors and not necessarily those of the Foundation.

Joseph Rowntree Foundation, The Homestead, 40 Water End, York YO30 6WP
Website: www.jrf.org.uk

Dimitris Ballas is a Lecturer at the Department of Geography, University of Sheffield.

David Rossiter is a Research Fellow at the School of Geography, University of Leeds.

Bethan Thomas is a Researcher at the Department of Geography, University of Sheffield.

Graham Clarke is a Professor at the School of Geography, University of Leeds.

Danny Dorling is a Professor at the Department of Geography, University of Sheffield.

First published 2005 by the Joseph Rowntree Foundation

ISBN 1 85935 265 0 (paperback)
ISBN 1 85935 266 9 (pdf: available at www.jrf.org.uk)

A CIP catalogue record for this report is available from the British Library.

Cover design by Adkins Design

Prepared and printed by:
York Publishing Services Ltd
64 Hallfield Road
Layerthorpe
York YO31 7ZQ
Tel: 01904 430033; Fax: 01904 430868; Website: www.yps-publishing.co.uk

Further copies of this report, or any other JRF publication, can be obtained either from the JRF website (www.jrf.org.uk/bookshop/) or from our distributor, York Publishing Services Ltd, at the above address.

Contents

Acknowledgements vii

Summary viii

PART I: BACKGROUND 1

1 Introduction 3
- Purpose 3
- Scope 3
- Perspective 4
- Approach 4
- Structure of the book 4

2 What is geographical microsimulation? 6
- Purpose 6
- Definition 6
- Types of microsimulation 8

3 The current state of geographical microsimulation 11

4 What spatial microsimulation can do and what it cannot! 14
- What can spatial microsimulation do? 14
- What spatial microsimulation cannot do! 14

PART II: CONCEPTUAL ISSUES 17

5 Spatial vs. aspatial microsimulation 19

6 Dynamic vs. static microsimulation and policy analysis 21
- Defining static microsimulation 21
- Defining dynamic microsimulation 23

7 Geographical approaches to analysing survey data 25
- Descriptive studies 25
- Mathematical and statistical models 26
- Strengths, weaknesses and limitations of past studies 28

8 Geographical data availability 29
- The census of population 29
- Other small area data sources 30
- Survey data 32

PART III: PRACTICAL APPLICATION 37

9 How to add geography to national survey data 39
Reweighting survey household records 39

10 Updating small area population statistics 44

11 Projecting small area microdata into the future 61
Reweighting the BHPS to fit small area population statistics for York in 1991–2021 61
Strengths and limitations 65

12 Projecting detailed survey data into the future 66
Background 66
The creation of GHOSTs 67
A regression framework for continuous variables 82

13 Assessing simulation outputs and policy relevance 83
Validating the simulated outputs with the use of census data 83
Validating the small area population projections 88

14 Examples of how policy analysis can be performed 91
Classifying the simulated households 91
Exploring the living standards of different household classes 94
Estimating policy impacts 100
Estimating the impact of welfare reforms introduced in April 2003 108
Strengths and limitations 111

15 Summary and conclusions 116

Notes 119

References 120

Appendix: A regression framework for continuous and binary variables 128

Acknowledgements

The authors are grateful to the Joseph Rowntree Foundation for funding this research. They are particularly grateful for the help and support of Danielle Walker, who also chaired the Advisory Group Meetings. They are also grateful to Ben Anderson (BTExact/University of Essex) and to the Foundation's external Advisory Group that provided many valuable comments, as well as informal advice between meetings: Roger Burrows, David Darton, Jane Falkingham, Trevor Huddlestone, Alan Jackson, Heather Joshi, Ade Kearns, Sarah Lowe, Howard Reed, Ludi Simpson, Holly Sutherland and Jane Todorovic. Thanks are also due to Heather Eyre, who worked on this project during its first stages, and to Bronwen Dorling for proofreading parts of this book.

The Census Small Area Statistics are provided through the Census Dissemination Unit of the University of Manchester, with the support of the ESRC/JISC/DENI 1991 Census of Population Programme. All census data reported in this paper are Crown Copyright. The BHPS data were obtained from the UK Data Archive (University of Essex). All digitised boundary data are Crown copyright.

The authors would also like to thank Alison Elks, Graham Allsop and Paul Coles for their help and their comments on the illustrations. Thanks are also due to Amanda Hancock and Caroline Hall for all their editorial and design work and to Lyn Hedges for her copy-editing work.

Summary

This book presents the results of a research project, which aimed to create a simulation model that can be used for the estimation of the spatial impacts of social policies, as well as their socio-economic impact. It is suggested that there is a need for spatial analysis of national social policies. The book aims to encourage the incorporation of geographical insights into traditional evaluation studies. In particular, it seeks to encourage the use of geographical data and methods by non-geographers who are involved in the analysis and evaluation of social policies.

The simulation methodology that is presented in this book builds on past work in the area of microsimulation. Microsimulation is a technique that has long been established in economics and has been widely used by governments around the world for the analysis of redistributive policies and budget changes. Nevertheless, there have been very few examples of extending these simulation models to enable the estimation of geographical impacts of policies. The book identifies the main reasons for this paucity of geographical microsimulation work and discusses the conceptual and practical issues of microsimulation, highlighting the differences between spatial and aspatial microsimulation models. It also explains the differences between static and dynamic microsimulation. The literature on geographical microsimulation is then briefly reviewed. This is followed by a brief review of other methods that have been used for the geographical analysis of survey data, as well as other data.

One of the most important features of this book is that it explains how a geographical microsimulation model can be built. One of the aims of the research project underlying the book was to develop a technique that would be simple enough to be used by any social scientist with a basic understanding of quantitative research methods. Therefore the model is based on a relatively simple method, which is explained with an attempt to keep mathematical and statistical jargon to a minimum. The method uses data that are publicly available and it is capable of generating small area populations at different points in time. More sophisticated enhancements to the model are described as well as demonstrations of how the method can be validated and used for the geographical analysis of social policies. It is hoped that the work presented here will promote more convergence of the methods used by economists, geographers and other social scientists who work in this field.

Part I
Background

1 Introduction

Purpose

This book is the main output of an 18-month research project, which aimed to construct prototype computer simulation models that can be used for geographical policy analysis. This project was based on a construction of two pilot spatial microsimulation models for the City of York and for Wales. This book argues the case for geographical analysis of national social policies and uses simulation outputs for Wales and York to illustrate the policy relevance and usefulness of the spatial microsimulation method.

The need for geographical analysis of national social policies is rarely acknowledged, as the main focus of policy evaluation strategies remains the likely impact on individuals and households irrespective of where they live. The report presented here seeks to highlight the importance of geography in policy analysis: it is argued that there is a need to estimate the spatial impacts of social policies, as well as their *socio-economic* impact.

The book is aimed at social scientists who are involved in the analysis and evaluation of social policies. It is hoped that it will encourage the incorporation of geographical insights into traditional evaluation studies. In particular, the book seeks to encourage the use of geographical data and methods by non-geographers who are involved in the analysis and evaluation of social policies. It also seeks to remind economic and social geographers that all government policies have spatial effects and to highlight the importance of looking at the geographical impacts of non-spatial policies. The book assumes a basic understanding of statistics and mathematics and of policy evaluation approaches. The relevant data sources and methods are briefly introduced in the book, with an attempt to keep statistical and mathematical jargon to a minimum.

Scope

The methods presented here are based on a population simulation approach called spatial microsimulation. Microsimulation is a technique that has been broadly developed and used by economists over the last 40 years.[1] The results of microsimulation models are widely quoted in the media when covering the possible impact of government budget changes on different types of households. However, as argued in this book, traditional economic microsimulation models do not take geography into account. Microsimulation is defined here as a method to construct small area population microdata for one point in time and then to update these microdata. The definition of microsimulation used here is slightly different from that given by economists who have been involved with building statistically and mathematically robust microsimulation models, which, however, do not take geography into account. The focus here is to shed light onto the geographical impacts of policies and to provide small area socio-economic information that can be used for the spatial analysis of policies, as well as the inter-household distributional effects (traditionally the main concern of economists and social policy practitioners).

The main goal is not to demonstrate how microsimulation models can be built and used to analyse policy impacts (this has been nicely done elsewhere, see for instance Redmond *et al.*, 1998), but to show how geographical models can be built and used to analyse the geographical impacts of social policies. A simple method of creating the necessary geographical datasets that can be used for policy analysis is presented. One of the aims was to develop a technique that would be simple enough to be used by any social scientist with a basic understanding of quantitative research methods. By aiming for simplicity, it was felt that input in this area would be much more valuable and that the work presented here would encourage economists and geographers to work more in this field.

Perspective

The work presented here looks at national policies from a geographical perspective. Historically, social scientists have not attempted to look at the geographical dimensions of national social policies. There is an increasing need to investigate the geographical implications of national policies, current trends in socio-economic polarisation and inequalities between and within cities. The method presented here can be employed to monitor and project trends in socio-economic polarisation, and inequalities if past trends were to continue. It can be used to address a series of important policy questions from a geographical perspective.

There is a wealth of geographical data from the census of the UK population that is currently underused. There is also an abundance of survey data that can be used, as is demonstrated in this book, in combination with census data to produce synthetic populations.

Approach

The work presented here aims to provide a simulation tool that can be used by any social scientist with an interest in spatial issues. A theoretical discussion of the microsimulation idea is presented and a simple method that can be used to create small area population microdatasets is suggested. The method presented uses small area data from past censuses of the British population in order to estimate small area data for 2001, 2011 and 2021. In particular, the method uses data from the censuses of 1971, 1981 and 1991 to estimate future trends in car ownership, class composition, demography, employment, household types and tenure at the small area level. These projected datasets are then used in combination with national survey data in order to simulate the living standards of different types of households in 1991, 2001, 2011 and 2021. The method presented here enables the exploration of a wide range of statistics at various geographical scales. These statistics include household earned income, health, labour market status and educational qualifications, which can in turn be used to provide estimates of trends in poverty and socio-economic polarisation at various geographical scales. It is argued that the generated geographical information is extremely policy relevant and it is shown how it can be used to analyse the spatial effects of national social policies.

Structure of the book

Chapter 2 provides a definition of geographical microsimulation and outlines alternative approaches to estimating and updating small area microdata.

Chapter 3 provides a brief review of the literature in geographical microsimulation and identifies the main related issues and challenges.

Chapter 4 gives a brief overview of the capabilities, strengths and weaknesses of spatial microsimulation.

Part II (Chapters 5–8) explores the various conceptual issues that are associated with the technique. In particular, it compares spatial and aspatial microsimulation and explores the reasons why most microsimulation models to date have disregarded spatial issues (Chapter 5). Further, it explains the differences between static and dynamic microsimulation models (Chapter 6). It also reviews the methods geographers traditionally used to analyse survey data (Chapter 7) and it then presents the geographical datasets and discusses their usefulness (Chapter 8).

Part III (Chapters 9–15) shows how a spatial microsimulation model can be built and used for policy analysis. In particular, Chapter 9 presents a simple methodology of reweighting survey data records so that they fit geographical cross-tabulations. Chapter 10 discusses a method of projecting small area data into the future. Chapter 11 shows how the method described in Chapter 8 can be applied to provide 'future' small area microdata with an application in the City of York. Chapter 12 shows how the latter data can be further updated, using a novel spatial household modelling method called GHOSTs (Generic Households Spaces through Time) as well as a regression modelling framework. Chapter 13 argues that the usefulness of any spatial simulation model depends on its ability to make simulation output data replicate what happens in the 'real world' and to project from there any likely distributional effects of changes in policy. It then describes how the modelling method presented here can be validated, using simulated and actual data in York and Wales. Chapter 14 shows how the method and the data that it produces can be used for policy analysis. Finally, Chapter 15 offers some concluding comments.

2 What is geographical microsimulation?

Purpose

The purpose of geographical microsimulation is to inform decisions about the spatial as well as the *socio-economic* impacts of policy decisions. All government policies have a geographical impact, irrespective of whether they are targeted to particular regions or small areas. Area-based policies have a geographical impact by definition and there is a wide range of evaluation methods that have been developed and used to analyse the effects of these policies. However, there has been very limited analysis of the spatial impacts of policies that were not designed to have a geographical impact. All policies have a spatial dimension, which becomes very important when compared to their area-based counterparts. Geographical microsimulation can be used to estimate the geographical impacts of national policies and inform decisions on the revision of these policies on the basis of their likely spatial as well as socio-economic distributional effects.

Definition

Microsimulation can be defined as a methodology that is concerned with the creation of large-scale population microdatasets and with the analysis of policy impacts at the microlevel. In particular, microsimulation methods aim to examine changes in the life of individuals within households and to analyse the impact of government policy changes for each individual and each household (Hancock and Sutherland, 1992; Harding, 1996; Mitton *et al.*, 2000). Nevertheless, there are relatively few examples of spatial models that build on traditional economic microsimulation frameworks by adding a geographical dimension. Geographical microsimulation techniques involve the merging of census and survey data to simulate a population of individuals within households (for different geographical units), whose characteristics are as close to the real population as it is possible to estimate (Clarke, 1996; Ballas, 2001). Geographical microsimulation models simulate virtual populations in given geographical areas, so that the characteristics of these populations are as close as possible to their 'real-world' counterparts. Microsimulation models are concerned with the creation of large-scale datasets estimating the attributes of individuals within households and are used to analyse policy impacts on these microunits (hence microsimulation; Orcutt *et al*., 1986; Birkin and Clarke, 1995; Clarke, 1996). Microsimulation can also be used to simulate entities other than individuals or households (e.g. the behaviour of firms). However, the material presented here refers to the spatial microsimulation of population. One of the major advantages of microsimulation is that it can be a substitute for conducting detailed surveys.

The microsimulation method typically involves four major procedures:

1. the construction of a microdataset from samples and surveys: in our case, a detailed representative sample of households in Britain (see Box 1).
2. sampling from this dataset to 'create' a microlevel population for individuals for small areas who match the known data on those areas
3. static *what-if* simulations (see Box 2), in which the impacts of alternative policy scenarios on the population are estimated:

for instance, if there had been no poll tax in 1991, which communities would have benefited most and which would have had to have paid more tax in other forms?

4 dynamic modelling (see Box 3), to update a basic microdataset and future-oriented *what-if* simulations: for instance, if the current Government had raised income taxes in 1997, what would the redistributive effects have been between different socio-economic groups and between central cities and their suburbs by 2007?

Box 1 How population small area microdata can be constructed

Integer reweighting: sample from a microdataset to fill in small area data

Integer reweighting involves the reweighting of an existing microdata sample (which is only available at coarse levels of geography), so that it would fit small area population statistics tables. For instance, an existing microdataset such as the British Household Panel Survey (BHPS) can be reweighted to 'populate' small areas. The BHPS provides a detailed record for a sample of households and all of their occupants. Reweighting methods aim to sample from all the microdata records to find the set of household records that best matches the population described in the Small Area Statistics or Census Area Statistics tables for the small area under study. First, a series of small area tables (e.g. from the census or other sources) that describe the small area of interest must be selected. For example, a reweighting method would sample from the BHPS to find a suitable combination of households that would fit the tables shown in Table 1 in two areas.

Table 1 Small area tables

Small area table 1 (household type)	Small area table 2 (economic activity of household head)	Small area table 3 (tenure status)
Area 1	*Area 1*	*Area 1*
60 married-couple households	80 employed/self-employed	60 owner-occupier
20 single-person households	10 unemployed	20 local authority or housing association
20 other	10 other	20 rented privately
Area 2	*Area 2*	*Area 2*
40 married-couple households	60 employed/self-employed	60 owner-occupier
20 single-person households	20 unemployed	20 local authority or housing association
40 other	20 other	20 rented privately

The task would be to select the records of the BHPS microdata that best match these tables. However, there are a vast number of possible sets of households that can be drawn from the BHPS sample. There is a wide range of techniques that can be employed to find a set that fits the target tables well.

Box 2 How 'what-if' analysis can be performed

'What would have happened if' examples
Assuming the British synthetic population is estimated for all the years up to 2001, it might be possible to estimate what would have happened if a specific policy had or had not been implemented, or if certain aspects of the British society were different. For instance, following a previous US study (Caldwell and Keister, 1996), if it is assumed that one of the findings of the baseline 1991–2001 simulation is that ethnicity plays an important role in social dynamics – in particular, if it is assumed that, according to the simulation findings, the probability of members of some 'minority ethnic households' improving their quality of life and getting better jobs is significantly lower than the respective probability of the rest of the population – it would be possible to re-run the 1991–2001 simulation on the basis that this probability is equal for all ethnic groups. The output of this alternative simulation would demonstrate what the distribution of income and wealth would be if race played no role. Respectively, it would be possible to simulate what the population of Britain would have been if social policy strategies had been different.

'What will happen if' examples
The initial simulation of the future population of Britain could be based on population projections from other sources (such as ONS surveys) and on the assumption that the trends in the changes in society to 2021 are similar to that of the previous decade. However, alternative projections could also be provided on the basis of *hypothetical progressive social policy schemes*. One issue to focus on is the eradication of child poverty. For example, a microsimulation model can be capable of estimating the degree of child poverty eradication within the next 20 years under different policies and assumptions, such as the onset of a major recession or a redistribution of wealth, and would provide projections in order to suggest where current strategies are failing to eradicate child poverty within a generation. A relevant example of microsimulation models that aim at estimating the impact of government policies on child poverty is the work of Sutherland and Piachaud (2001) and Sutherland *et al.* (2003).

Types of microsimulation

Microsimulation models can be distinguished between various types. For instance, there are static models that are based on simple snapshots of the current circumstances of a sample of the population at any one time and dynamic models that vary or age the attributes of each microunit in a sample to build up a synthetic longitudinal database describing the sample members' lifetimes into the future (Mertz, 1991). Further, microsimulation models can become *geographical* when spatial information about the simulated entities is available (or estimated).

As noted above, the first step in *geographical microsimulation* is the construction of a small area microdataset. Although the number of published population microdatasets is increasing around the world, there are still many cases in which this kind of data is not available at the desired spatial scale (Birkin and Clarke, 1995; Clarke, 1996). In the case of the United Kingdom, a very important source of microdata is the Sample of Anonymised Records (SARs). These are samples of individual census records that are anonymised in various ways ensuring that there is no breach of the confidentiality of the census and that no individual can be identified from the data (Marsh, 1993; Middleton, 1995). However, the spatial scale at which these datasets are released is, at best, the regional or district scale. In cases where official microdata are not available at the desired spatial scale, then this type of data can be estimated with the use of existing datasets and of a variety of techniques ranging from iterative proportional fitting methods to linear programming and complex combinatorial optimisation methods (Williamson *et al*., 1998; Ballas and Clarke, 2000; Ballas, 2001). These techniques make up stage 1 of the microsimulation procedure outlined above. It can thus be argued that static spatial microsimulation models can be distinguished between the following types:

- synthetic probabilistic reconstruction models, which involve the use of random sampling
- reweighting probabilistic approaches, which typically reweight an existing national microdataset to fit a geographical area description on the basis of random sampling and optimisation techniques
- reweighting deterministic approaches, which reweight a non-geographical population microdataset to fit small area descriptions, but *without* the use of random sampling procedures.

In turn, dynamic spatial microsimulation models can be distinguished between two types:

- probabilistic dynamic models, which use event probabilities to project each individual in the simulated database into the future
- *implicitly* dynamic models, which use independent small area projections and then apply the static simulation methodologies to create small area microdata *statically* (see Box 3).

Box 3 Dynamic models: an example

The dynamic probabilistic modelling approach to dynamic microsimulation involves behavioural modelling, which, as will be argued below, is an extremely difficult task. One of the inherent difficulties of such a task is to determine the interdependencies between individual attributes and events. For instance, the probabilities of an individual participating in the labour force may be conditional on family status (e.g. having children). However, it may also be argued that family status depends on labour market status. As Falkingham and Lessof (1992) put it:

(Continued overleaf)

… while a woman's labour force status can depend on the number of children she has and on her marital status, it cannot also influence the probability of the woman having a child in any year. The ordering of the modules necessarily involves making assumptions about the direction of causality in relationships between variables.
(Falkingham and Lessof, 1992, p. 9)

An additional difficulty associated with dynamic spatial microsimulation models is the lack of sufficient geographical data that would enable the simulation of interactions such as migration flows between areas (e.g. there are no microdata on migration that would enable a reasonably accurate simulation of migration into the future). Due to the lack of suitable data there have been very few examples of spatial microsimulation of events such as migration (e.g. see Swan, 2000; Ballas *et al*., 2001).

An alternative and much simpler way of generating estimates of future population microdata is by using the integer reweighting method described in Box 1 to find suitable combinations of households that match projected small area tables. Such an approach is adopted in the context of this book and is presented in Part III.

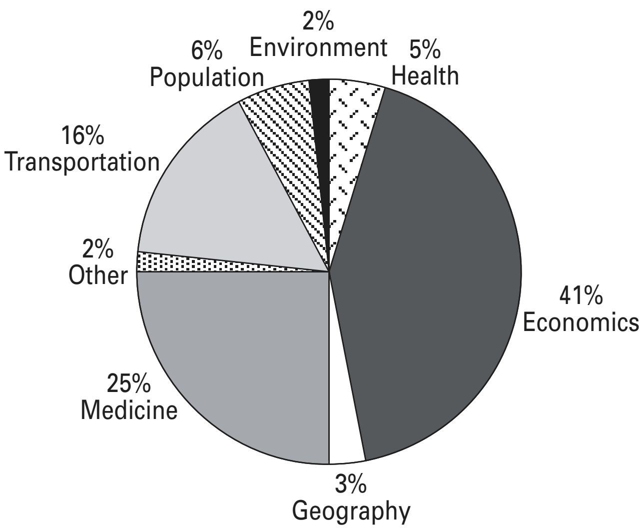
3 The current state of geographical microsimulation

As mentioned above, microsimulation has been mainly developed and used by economists and there have been relatively few examples of geographical microsimulation. Figure 1 shows the results of a basic keyword search in the *Sciencedirect* academic journal database, searching the word 'microsimulation' in the titles or abstracts of papers in the last 30 years. As can be seen, the majority of the papers were in economics (41 per cent) with very few papers in geography (3 per cent). There is also a relatively high number of microsimulation applications in medicine. However, these are applications of a different nature, as their main focus is the effectiveness of medicines (e.g. simulating the impact of medicines, etc.).

It is interesting to note that, during the 1990s, many studies of British society made use of some form of microsimulation. The technique has grown in use recently as the computer-intensive nature of the work has become less of a handicap. However, many of these studies did not recognise that the techniques they were using fell into this category and they often had to 're-invent the wheel' in their work. Furthermore, these studies were usually unique. A single issue would be identified; the work done over a couple of years to address the issue; and the results published – often after the salience of the issue had diminished and the data appeared to be further out of date.

It has long been argued that, in most cases, the microsimulation models developed so far do not take spatial scale into account (Birkin *et al*., 1996). Among the notable exceptions has been the work conducted at Leeds University (Birkin and Clarke, 1988), which used static spatial microsimulation techniques to generate a synthetic microdatabase for the Leeds Metropolitan District. Moreover, Birkin and Clarke (1989) used this SYNTHESIS model to generate incomes for individuals. Williamson

Figure 1 Distribution of microsimulation academic studies in the period 1967–2003 (source: http://www.sciencedirect.com/; accessed 15 October 2003)



(1992) developed a model for the spatial analysis of community care policies for older people, while Ballas *et al.* (1999) and Ballas and Clarke (2000) argued and demonstrated the case for a geographical microsimulation approach to local labour market analysis and developed a spatial microsimulation model for the Leeds local labour market. In addition, Ballas and Clarke (2001a) adopted a spatial microsimulation approach to the analysis of the spatial impacts of social policies. In particular, they used spatial microsimulation methodologies to perform static *what-if* policy analysis and identify which types of households and which geographical areas would be affected the most under different social policy scenarios. In another study (Ballas and Clarke, 2001b), static spatial microsimulation was used to evaluate the socio-economic impact of a plant closure in east Leeds.

The studies mentioned above were explicitly labelled as microsimulation, but it is important to note that even calculating simple social statistics such as life expectancy is a form of microsimulation. In life expectancy calculations, probabilities derived from people who have already died are applied to estimate the likely lifespans of those currently alive. Recent work for the Joseph Rowntree Foundation (Mitchell *et al*., 2000) incorporated a simple form of microsimulation to ask what would happen were 'Britain more equal' in terms of health. Another example is the research conducted by Dorling (1994) who considered how housing market debt might have changed under different scenarios using a simulation based on a million households (Dorling, 1994). Work funded by ESRC used similarly large datasets to establish the degree of social polarisation taking place in Britain (Dorling and Woodward, 1996), to examine voting trends (Johnston *et al*., 1998) and the long-term effects of migration (Brimblecombe *et al*., 2000). All of these studies were limited to particular issues and did not highlight the fact that the methods being used involved microsimulation of one kind or another.

Microsimulation still has to gain credibility among the social science community in general and social policy researchers in particular. One of the aims of this book is to remedy this situation. It should also be noted that there is currently a major challenge to build on the work described above in order to project the population into the future to predict what would happen under different macroeconomic, microeconomic and social policy scenarios. This will enable an evaluation of the short- and long-term impacts that various government policies are likely to have on different segments of British society and different geographical areas.

Although no serious attempts have been made to simulate spatially and dynamically the population of Britain, there are numerous examples of successful dynamic microsimulation models being used outside Britain. In particular, Caldwell and Keister (1996) and Caldwell *et al*. (1998) describe CORSIM, which is a dynamic spatial microsimulation model under development at Cornell University since 1986. CORSIM has been used to model wealth distribution in the United States over the historical period 1960–95 and to forecast wealth distribution over the future. Another example of dynamic spatial microsimulation is the ongoing work in the Netherlands described by Hooimeijer (1996), which adopts a spatial microsimulation approach to analyse simultaneously the linkages between supply and demand in the housing market and in the labour market. In the context of this work, the spatial mobility of households is modelled in three different time sets (daily commuting, relocation, lifetime mobility). The fundamental characteristic of this methodology is the life-course approach to the behaviour of households (Hooimeijer, 1996).

A more recent example of comprehensive dynamic spatial microsimulation modelling is the work of the Spatial Modelling Centre (SMC)[1] in Sweden (Holm *et al*., 1996; Vencatasawmy *et al*., 1999; Swan, 2000). The SMC built on previous microsimulation modelling efforts (Holm *et al*., 1996) and constructed TOPSIM (Total Population Simulation Models) and *SVERIGE* (System for Visualising Economic and Regional Influences Governing the Environment), which simulates the entire population of Sweden. In addition, *SVERIGE* is aimed at studying the spatial consequences of various national, regional and local-level public policies. The database used for this model comprises longitudinal socio-economic information on every resident of Sweden for the years 1985 to 1995.

There are very few examples of dynamic spatial microsimulation in the UK. For instance, Clarke (1986) articulated a spatial microsimulation methodology for the modelling of demographic processes and household dynamics. In addition, Duley *et al*. (1988) and Duley (1989) constructed a demographic microsimulation model for updating population data at the small area level between censuses.

Although there are no other attempts to build dynamic spatial microsimulation models in the UK, there is a great deal of dynamic microsimulation work conducted at the *national level*. The first model of this kind is LIFEMOD, which is a dynamic cohort microsimulation model, simulating the life histories of a cohort of 2,000 males and 2,000 females (Falkingham and Lessof, 1992; Falkingham *et al*., 1995). Each individual 'experiences' major life events such as schooling, marriage, childbirth, children leaving home and employment. The LIFEMOD model has been used to estimate the effects of the welfare state over the life-cycle of individuals (Falkingham and Hills, 1995a), as well as to estimate the degree to which income is redistributed between people over time, or across the life cycle (Falkingham and Hills, 1995b). It has also been used to investigate financing options for higher education (Glennerster *et al*., 1995) and the dynamics of lone parenthood (Evandrou and Falkingham, 1995). Further, LIFEMOD has been used to explore the lifetime distribution of health needs and use of health services (Propper, 1995). Another example of a UK national dynamic microsimulation model is PENSIM (Hancock *et al*., 1992), which aimed to study the influences of policy change on the income distribution of pensioners up to 2030. A more recent example of dynamic microsimulation modelling in Britain is the work of Hancock (2000), who simulated the contribution that older people would make towards the cost of care both at present and in 15 years' time.

4 What spatial microsimulation can do and what it cannot!

The research presented here demonstrates how it is possible to build a geographical simulation model that could be capable of simulating survey-style data at different geographical scales. The following chapters discuss various conceptual issues pertaining to microsimulation in some detail. However, before examining these issues, it is useful to briefly outline the strengths and weaknesses of geographical microsimulation frameworks.

What can spatial microsimulation do?

Spatial microsimulation frameworks can be used to paint a picture of the possible (or most probable) life of households of a city or region at various geographical scales. For instance, this book shows how spatial microsimulation can be used to paint a picture of York (Chapter 13), in which there is a trend for polarisation that appears to be determined largely by educational qualifications, with more children being in relative poverty in the future, as their parents tend not to be so well qualified (and hence rewarded), if current trends continue. Further, as argued in Chapters 13 and 14, spatial microsimulation is suitable for the estimation of variables such as household income at the small area level, where it is the only legal way of estimating such variables. Such estimations can provide helpful insights into the analysis of spatial and socio-economic polarisation within cities. Spatial microsimulation can also be used to paint a picture of the life of households of different income categories. In particular, the simulation outputs presented in this book are very similar to complex survey outputs and qualitative research findings. Nevertheless, it should be stressed that spatial microsimulation outputs are model estimates (the reliability of which depends on a wide range of factors) and not findings.

Spatial microsimulation has been claimed to be useful in modelling the socio-economic and spatial effects of policy change. In particular, in the context of this book (Chapter 13), we investigate the effects of policy changes over the last ten years (including the 2003 new Tax and Working Credits). The analysis suggests that some of these newly introduced policies and measures would have a dramatic effect on reducing child poverty if full uptake were assumed.

Overall, spatial microsimulation frameworks such as the one presented in this book can be used to provide useful 'most probable' information on socio-economic trends, as well as on the possible outcome of policy reforms, at different geographical scales. Further, spatial microsimulation models can be used as tools that could aid policy makers to think more geographically about the potential effects of policy options they may consider.

What spatial microsimulation cannot do!

Before discussing the conceptual issues underpinning spatial microsimulation, it should be noted that most spatial microsimulation models (including the one presented in this book) are unsuitable for the prediction of variables that are affected considerably by external and localised factors, such as transport networks and public transport services, or the presence of a disproportionately large university or a single major employer in the region. Further, most spatial microsimulation models cannot be used to analyse the longer-term behavioural responses to policy changes. For instance, the model presented in this book cannot be used, in its current form, to predict how many unemployed individuals would decide, if they could, to enter the labour market as a result of increases in the Minimum Wage, or welfare-to-work policies such as the Working Families' Tax Credit.

The remainder of this book discusses microsimulation and spatial microsimulation in some detail, before presenting an example of how a spatial microsimulation model can be built and used for the geographical analysis of social policies.

Part II
Conceptual issues

5 Spatial vs. aspatial microsimulation

One of the main distinctions, which is rarely noted in the microsimulation literature, is that between spatial and aspatial microsimulation. As noted above, microsimulation has a long history in economics, which led to the acceptance of the microsimulation method as a standard tool for the evaluation of economic and social policy, and in the analysis of tax-benefit options and in other areas of public policy (Hancock and Sutherland, 1992; Harding, 1996; Mitton *et al*., 2000). The standard non-geographical microsimulation models have been built on a very good basis that was formed during the course of systematic research by economists over the last 40 years. During that period, geography has been persistently ignored by microsimulation researchers and there are several reasons for this.

- *Lack of good quality geographical data*: there were very few sources of geographical socio-economic data. Even today, there are no small area population microdata, which is the standard dataset used by economic microsimulation models.
- *Computational intensity*: the incorporation of geography into standard microsimulation models increases significantly the computational demand.
- Concerns with *simulation accuracy*.
- Belief that *geography is not important*.
- Unfamiliarity with geographical data and methods.

Some of these problems have recently been tackled because of an accelerating growth in the volume, variety, power and sophistication of the computer-based tools and methods available to support urban and regional analysis and policy making. Developments in hardware and software systems have enabled significant advances to be made in the storage, retrieval, processing and presentation of spatially referenced data (Scholten and Stillwell, 1990). There has also been significant progress in the development of Geographical Information Systems (GIS) for socio-economic applications (see, for instance, Scholten and Stillwell, 1990; Martin, 1996; Longley *et al*., 1999; Stillwell and Scholten, 2001). Further, there has been an increasing availability of a wide range of new geographical data sources in both the public and private sectors, and an increased power and portability of personal computers bringing high quality computer graphics and presentation facilities (Bertuglia *et al*., 1994; Birkin *et al*., 1996).

In this new environment, there have been many spatial models that have shed new light on patterns and flows within cities and regions. These models, when combined with relevant performance indicators, have been very useful in measuring the quality of life for residents in different localities (Bertuglia *et al*., 1994; Clarke and Wilson, 1994). However, relatively little is known about the *interdependencies* between household structure or type and their lifestyles, including the events they routinely participate in and hence their ability to raise and spend various types of income and wealth. The modelling of interdependencies requires a different level of urban and regional system representation.

In this context, spatial microsimulation offers a potentially powerful framework for policy analysis. This book aims to encourage the closer collaboration of economists and geographers in order to make the most of the geographical simulation and aspatial microsimulation methods. It is possible to link traditional microsimulation that has been well developed by economists to geographical methods and data. This synergy would enable the better co-ordination of both area-based and national policies. A synergy would enable researchers to move on from traditional deprivation indices to better methods of assessing poverty, such as households below half of median household income by area.

Further, spatial microsimulation models can be used to paint a picture of a city or region by simulating detailed social surveys, which would be too expensive to conduct. In addition, spatial microsimulation models can be used to provide useful information on socio-economic trends, as well as on the possible outcome of policy reforms, at different geographical scales.

6 Dynamic vs. static microsimulation and policy analysis

The most common classification of microsimulation models is to divide them between static and dynamic microsimulation. As described in Chapter 2, static microsimulation models are based on simple snapshots of the current circumstances of a sample of the population at any one time, whereas dynamic models are based on synthetic longitudinal databases describing the sample members' lifetimes. In this chapter, these differences are discussed from a geographical perspective and it is argued that there is a possibility for wider definitions of static and dynamic models when a geographical dimension is added. There are relatively few examples of spatial models that build on traditional economic microsimulation frameworks by adding a geographical dimension.

Defining static microsimulation

Static microsimulation involves the analysis of a population microdataset at one point in time for policy analysis. For instance, economists have been involved in the development of static microsimulation models that are capable of answering questions such as the following.

- What would be the impact of Working Families' Tax Credit on different types of households and individuals in its initial year of application?
- What would be the redistributional impacts of the government budget changes at one point in time?
- What would be the impacts of alternative policies on child poverty?
- How could new tax credits be funded through taxation?

The advantage of aspatial microsimulation models that have been used by economists is that they rely on existing good quality microdatasets (such as the Family Expenditure Survey and the Family Resources Survey) and they focus on using these datasets. Recent examples of aspatial microsimulation include the work of Sutherland and Piachaud (2001) and Sutherland *et al.* (2003) who used a microsimulation model to estimate the impact of government policies on poverty in Britain since 1997.

Adding spatial detail to traditional microsimulation involves *creating* a microdataset, as well as using it. There are very few sources of geographically detailed microdatasets, so there is a need to create these datasets using static geographical microsimulation techniques. It is interesting to note at this stage that economists would probably not use the term microsimulation for the creation of a microdataset, whereas, in geography, most of the effort in microsimulation models has been in constructing good quality geographically disaggregated population microdata. Geographical microsimulation techniques involve the merging of census and survey data to simulate a population of individuals within households (for different geographical units), whose characteristics are as close to the real population as it is possible to estimate. They can then be used to answer questions such as the following.

- How does the quality of life of individuals and households vary across different regions, cities and neighbourhoods?
- What are the interdependencies of household characteristics with geographical factors such as the presence of hospitals, community centres, schools, etc. in an area?
- To perform static *what-if* scenario analysis: e.g. answer questions such as 'what would happen to personal accessibilities if the patterns of service provision change?'
- What would be the geographical impact of national social policies on personal incomes and how effective would it be compared with an alternative area-based policy?

Questions such as the last one have become increasingly important given that a large (maybe the largest) part of the social and economic needs of residents in deprived areas are not met by area-based policies but by national social policies. As Edwards (1995) put it:

> *… by far the greatest part of the social and economic needs of inner city residents will not be met by urban specific policies but by mainline housing, health, income support and education provision.*
> (Edwards, 1995, p. 701)

There is a need for urban and regional planning based on social policies as well as area-based policies. It has often been argued that the allocation of extra resources for areas deemed to be deprived may only ever lead to cosmetic improvements. It would therefore be very useful to explore the spatial outcomes of policies that have not been designed to address geographical issues:

> *… inner city residents and in particular the urban deprived do receive most of their welfare by way of mainstream programmes and we know precious little about the effectiveness of such programmes either in targeting the deprived, spatially or otherwise, or in providing for the (sometime) multiplicity of needs or the different or additional needs that may be found within individual households.*
> (Edwards, 1995, p. 711)

Further, one of the problems that social policies aim to tackle is poverty. However, as McCormick and Philo (1995) point out, much poverty in the UK is hidden in the sense that *poor* people and localities are largely invisible. Further, they argue that poverty in these localities is not only the result of economic decline reflected as a shift in demand for specific labour market skills, but also the cause of the decline. In particular, it seems that there is a vicious circle of poverty in these areas:

> *Reducing the purchasing power of low-income households – which are likely to spend money locally rather than saving it, spending it elsewhere or using it on expensive imported goods – damages the economies of the 'poor places'. A gradual process of uncoupling hence occurs between local and national economic trends, such that improvements in the latter may no longer feed through into any improvement in the former.*
> (McCormick and Philo, 1995, p. 11)

Changes in national social policies have different and major implications for household incomes in different areas. Therefore, there is a need to add geography to the already powerful analytical capabilities of the microsimulation models that have been developed by economists.

Defining dynamic microsimulation

Dynamic microsimulation involves forecasting past changes forward to produce the best possible estimate of an individual's circumstances in the future – were current trends to continue, or were they to change under different policy scenarios. Dynamic microsimulation typically involves the modelling of behavioural and second-order effects. This can be carried out on the basis of calculated probabilities for a series of event changes that occur during the lifetime of individuals:

> *During their lifetimes, the simulated individuals have to change their educational and employment status. They will enter school with different probabilities when they are between 14 and 20 years old, they will be employed in different jobs, lose their jobs, earn an income which depends on their type of job, and eventually retire with different probabilities depending on their ages.*
> (Gilbert and Troitzsch, 1999, p. 59)

However, behavioural modelling is an extremely difficult task and this is reflected in the relatively small number of dynamic microsimulation models to be found worldwide. It should also be noted that small area socio-economic forecasting is another very difficult task, which is also reflected in a relatively small number of attempts to carry out such a task (e.g. Cambridge Econometrics – see Barker *et al*., 2001).

Another aim of dynamic spatial microsimulation is the analysis of household and individual reactions and behavioural changes that may result from policy changes. This adds further to the complexity of the task:

> *… modelling the labour supply-response to policy changes within a model that only addresses the household sector raises the question of the supply of jobs and how this is affected, in the first place by the policy change and in the second place by the shift in labour supply. Similar problems arise in the detailed modelling of other household responses to policy changes.*
> (Redmond et al., 1998, p. 8)

The task becomes even more difficult when there are attempts to introduce geographical detail. Spatial dynamic microsimulation involves the behavioural modelling of individuals over time and at various geographical scales. It also involves the modelling of individual decisions that have a strong geographical element such as migration. The latter is dependent on a series of individual characteristics such as age, socio-economic background and tenure:

> *It is well known that mobility rates are substantially higher among renters than among homeowners. Similarly, the age structure of migrants to and from neighborhoods is likely to be quite different in a neighborhood comprised primarily of homeowners in comparison with a renter-dominated neighborhood.*
> (Rogerson and Plane, 1998, p. 1468)

Spatial dynamic microsimulation involves the modelling of different types of transitions on the basis of each individual's attributes and circumstances. This kind of analysis, which is related to the dynamics of urban systems, can be labelled *event modelling*. As is the case with the aspatial dynamic microsimulation, geographical dynamic microsimulation aims to analyse policy impacts over a period of time. One of the particular strengths of dynamic spatial microsimulation is that it can address policy questions from a geographical perspective. For instance, the opening of a new engineering plant would generate new jobs and consequently lead to increases in the disposable income of the individuals getting these jobs. The expenditure generated from such an *event* could generate second- and third-round expenditures triggered from an increase of other households' consumption. For instance, a change of employment status for 2,000 individuals, as a result of the plant opening, would alter their income circumstances and could trigger further *event* changes such as increase of household expenditure on groceries, change of retail location (moving to more expensive stores), etc. These new *events* could in turn lead to new developments, such as the opening of new supermarkets in an area with the subsequent creation of new jobs. Therefore, there is a challenge here to estimate not only the direct effects of regional investment but also the indirect and induced effects at the microscale. These effects are addressed by regional input-output models, which represent another branch of economics (see next chapter for a brief discussion of these models).

Nevertheless, one of the biggest problems associated with both spatial and non-spatial dynamic microsimulation is that they can be extremely complex and difficult to develop, implement and explain to policy practitioners who may be interested in using them. It has often been argued in the microsimulation literature that there is a need for transparency and simplicity in the construction of models. An alternative to the traditional comprehensive dynamic microsimulation models is to combine aggregate projection methods with the static microsimulation methods. This approach has been adopted in the context of this book and is explained in some detail in Part III.

7 Geographical approaches to analysing survey data

Geographers and regional scientists have traditionally been involved in building models that exploit the wealth of social survey data. This chapter gives some examples of the ways in which social scientists with an interest in geography have used survey data in their research.

Descriptive studies

One of the most obvious ways of analysing survey data in a geographical manner is by pure description of the datasets, and by mapping and exploring potential geographical patterns. A related area, where there have been numerous geographical studies that exploit survey data, has been the analysis of poverty at various geographical scales. For instance, Martin (1995) used income inequality statistics to demonstrate disparities between northern and southern regions, and he discussed the political and economic causes of these disparities. Moreover, McKendrick (1995) points out that a less obvious, but equally significant, division in income distribution is that between the Celtic nations of the UK (Scotland, Wales and Northern Ireland) and England.

At a smaller area level, Green (1996) argues for a spatial perspective on poverty and wealth, and presents selected evidence on changes and continuities in poverty and wealth between 1981 and 1991 at the spatial levels of *electoral ward, local authority district* and *local labour market area*. In addition, Goodwin (1995) examined the distribution of poverty at the intra-urban scale. He points out that one of the problems faced by those attempting to investigate urban poverty is the lack of any systematic small area level data on household income, wealth or living standards:

> *… an immediate problem faced by those attempting to assess the extent and the shape of urban poverty is the difficulty of obtaining reliable data on variables such as income, especially at those geographical scales which allow comparisons to be made within, as well as between, urban areas.*
> (Goodwin, 1995, pp. 66–7)

Further, he addresses the problems of the 'correct' geographical scale for the analysis of urban poverty and he argues that any conclusions that casually draw a link between certain types of urban locality, such as inner city areas and high concentrations of poverty, should be rejected. He then presents mapping scores of deprivation for Greater London wards, but also stresses the importance of the subjective dimensions of deprivation:

> *People experience these deprivations differently, and we should perhaps speak of their varying* experiences *of poverty. Different groups are affected in different ways, although some sections of society are more prone to poverty than are others.*
> (Goodwin, 1995, p. 78, emphasis in the original)

Another example of using survey data for the analysis of poverty is the work of Dorling and Tomaney (1995). Their analysis was based on the 'five great evils' of *want*, *ignorance*, *idleness*, *squalor* and *disease*, as defined by William Beveridge in the 1940s (Dorling and Tomaney, 1995). Using data from five different sources, they generated contemporary indicators for these 'evils' at the ward level for England and Wales, and produced cartograms reflecting the geographical complexity of poverty. Moreover, Dorling and Woodward (1996) investigated spatial inequalities and local social polarisation in Britain using data from the censuses of UK population for 1971, 1981 and 1991. In particular, they developed a methodology to harmonise the output of these censuses in order to allow the populations of over 10,000 electoral wards to be compared. They then investigated the degree to which various socio-economic population groups were spatially polarised across British society.

Further, Noble and Smith (1996) examined the spatial patterns of income and wealth in Oxford and Oldham at the intra-urban level, using data from the Housing Benefit/Council Tax Benefit (HB/CTB) systems at the individual claimant level in an anonymised form. They used these data to distinguish claimants in receipt of Income Support (IS) from those who are otherwise on low income and receive HB/CTB. They used the postcode of each individual record to assign it to its respective enumeration district (ED). They then constructed an index of low income and performed cluster analysis, comparing the intra-urban spatial patterns of Oxford and Oldham. This analysis showed that, although the overall proportions on benefit are similar, the geographical distributions at the ED level within the two towns are quite different. In a similar study, Noble *et al*. (2001) used a wide range of datasets to construct various measures of deprivation in Northern Ireland at the electoral ward level. A related study is the work of Connolly and Chisholm (1999) who presented a critique of the Index of Local Deprivation (ILD), which has been used by the UK Government for the targeting of resources to areas. They also made suggestions on how this index could be changed to provide a better basis for assessing relative need.

Mathematical and statistical models

The above brief discussion of the descriptive analysis of survey data in a geographical context is useful. However, there have been attempts to build on these analyses by constructing mathematical and statistical geographical models that can be used for a more sophisticated analysis of spatial socio-economic trends and for policy evaluation. Examples of this kind of work include the work of Curtis (1995) who investigated the link between health and poverty and examined these issues for different parts of the UK. In particular, she looked at illness and mortality differences in relation to various indicators of socio-economic status and for different age groups in different UK regions.

Further, Burrows and Rhodes (1998) used survey data in combination with census data to identify areas in Britain where residents were likely to display high levels of dissatisfaction with their neighbourhoods. In particular, they used data from the annual Survey of English Housing on residents' views of their area to construct a measure of area dissatisfaction and to examine what socio-economic variables best explain spatial variations in dissatisfaction. They also combined these variables with data from the 1991 UK census to generate estimates of the proportion of the population at different geographical levels who were likely to express high levels of dissatisfaction.

Also, Mitchell *et al*. (2000) built a model that investigated how the 'health gap' in Britain could be narrowed if different social policies were implemented. In particular, they estimated the impact of changes to the population in different areas of Britain under three different policy scenarios: *a modest redistribution of 'wealth', achieving 'full employment'* and *eradicating 'child poverty'*. They also estimated the combined effect that these policies would have on the populations of each British parliamentary constituency.

There is also a great deal of mathematical modelling work conducted by regional economists who have been involved in the construction of regional econometric and input-output models. In particular, there has been a lot of work on extending the econometric models that have traditionally been used by economists. Econometric methods and models have long been used for the measurement and evaluation of various economic policy impacts, mainly at the national level of the economy. Nevertheless, over the past few decades, econometricians have moved steadily from models of nations, and even multinational systems, towards lower levels of spatial disaggregation. A notable example of spatial econometric analysis is the work done by Cambridge Econometrics. This model has considerable data requirements. In particular, it is a time series, cross-section model distinguishing 49 industries and 68 categories of consumers' expenditure (Barker and Peterson, 1987). It requires a detailed knowledge of the structure of gross and intermediate output and of final demand in each region. However, such information is not available for the English regions. Cambridge Econometrics used a number of techniques to estimate this information. Barker *et al*. (2001) report on recent developments and applications of the Cambridge Econometrics model and present how the model was used to evaluate the effects of high sterling worth (as compared to other countries) on the economic performance of UK regions.

Similar approaches to building mathematical models on the basis of socio-economic survey data comprise the input-output and social accounting methods. These aim at modelling explicitly the input-output industry and household linkages that exist within a region. The input-output approach to modelling the economy has its roots in the work of Leontief (1951) on inter-industry economic analysis. As Armstrong and Taylor (1993) point out, the input-output modelling method is based on the simple but fundamental notion that the production of output requires inputs. The input-output modelling approach to the regional economy represents all the linkages in a transactions table or flow matrix (Isard, 1960, 1975; Armstrong and Taylor, 1993). In particular, the flow matrix records all the production flows that occur within the regional economy during a specific year. Many of these models have been used for the assessment of the impacts of policy changes or major investments or disinvestments in the local economies (for examples, see Jin and Wilson, 1993; Madden, 1993; Batey and Madden, 1999, 2001). In particular, input-output models can show three types of impact assessments. The first are direct impacts, which are the jobs generated by the new investment itself. Second, are the indirect impacts, which include the jobs created through the increased business for other business/industrial organisations generated by the new investment (inter-industry links). Third, there are the induced impacts. These are jobs generated as a result of additional household income and expenditure. These might be associated with the extra expenditure spent on retailing, transport, leisure, childcare, etc.

Strengths, weaknesses and limitations of past studies

The descriptive studies such as those briefly discussed in this chapter provide very useful pictures of the life of households across different geographical regions and localities. Descriptive studies are relatively straightforward to conduct and they can inform policy decisions. On the other hand, mathematical and statistical spatial models are potentially more powerful than purely descriptive studies. Further, they may add value to existing datasets by estimating new information that is not available by published sources (e.g. spatial econometric models may be used to estimate regional consumption levels under different scenarios). Nevertheless, they are much more difficult to build and implement, and they are based on assumptions that are not always realistic.

This book aims at providing a tool that can be used to enhance the potential of existing geographical methods. In particular, the method advocated in this book aims to add value to existing datasets by estimating new information. Nevertheless, the method presented here is relatively simple and easy to understand and implement. The next chapter discusses the geographical socio-economic and demographic datasets that are currently available in Britain, and suggests that there is a lot of potential to add value to these datasets by combining them with national survey data.

8 Geographical data availability

There are a number of geographical datasets that can be used for policy analysis. In this chapter, the main sources of geographical socio-economic and demographic information are described, in particular the datasets that become available every ten years from the census of population returns. Some other sources of geographical socio-economic information are also briefly described.

The census of population

The UK census of population has been and still remains the most authoritative social accounting of people and housing in Britain, and is a unique source of data for the social sciences (Dale, 1993; Rees *et al.*, 2002). The census records demographic and socio-economic information at a single point in time and is normally carried out every ten years. Census datasets describe the state of the whole national population and are extremely relevant for the analysis of a wide range of socio-economic issues and related policies. The topics covered by the census are determined, among other factors, by the necessity to preserve comparability over time and the need for timeliness (Dale, 1993). The most recent census in the UK was held in April 2001. Box 4 describes the questions asked by the 1991 census of the UK population, as well as the revisions and new questions in the 2001 census.

Box 4 UK census questions (Dixie and Dorling, 2002; Openshaw, 1995, p. 2)

Principal 1991 census questions

1 Household questions
- Type of accommodation
- Number of rooms

(Continued)

- Tenure
- Amenities
- Car and van ownership

2 Individual person questions
- Name
- Gender
- Date of birth
- Marital status
- Relationship in household
- Whereabouts on census night
- Usual address
- Term-time address
- Usual address one year ago
- Country of birth
- Ethnic group
- Long-term illness
- Whether working, retired, looking after house, etc.
- Hours worked per week
- Occupation
- Name and address of employer
- Address of place of work
- Daily journey to work
- Degree, professional and vocational qualification

New questions in the 2001 census
- General health
- Provision of unpaid care
- Time since last paid employment
- Size of employer's organisation
- Voluntary question on religion

2001 census question revisions
- Qualifications
- Relationship within the household
- Ethnic group
- Accommodation

As Openshaw (1995) points out, it may seem surprising that the 26 questions listed in Box 4 can generate a census dataset that is seemingly so complex to understand, handle and analyse. Further, as Dale *et al*. (2000) point out, the availability of census datasets can range from, at its most restrictive, summary tables with very limited detail on either individuals, households or areas, to, at the other extreme, a microdata file with complete population coverage and high level of detail on all variables. The UK census results are presented as cross-tabulations of the questions listed in Box 4 at different levels of geography, designed in a way that preserves confidentiality (Marsh, 1993; Openshaw, 1995). In particular, there are several cross-tabulations of census variables available at various levels of geographical detail, ranging from national level, down to the enumeration district (ED) level, which is an area of 200 households in average (Cole, 1993; Dale, 1993; Denham, 1993; Coombes, 1995; Rees, 1995). The main predefined sets of cross-tabulations of two or more census variables are the Local Base Statistics (LBS) and the Small Area Statistics (SAS) in 1991 and the Census Area Statistics (CAS) in 2001. The 1991 LBS comprised 99 tables for Great Britain containing approximately 20,000 statistical counts and are available down to electoral-ward level in England and Wales and postal sector in Scotland (Cole, 1993; Rees, 1995). Further, the 1991 SAS are an abbreviated version of the LBS and consist of 86 tables for Great Britain, containing approximately 9,000 statistical counts and are available down to ED level in England and Wales and output area (OA) in Scotland (for more details, see Cole, 1993 and the appendix in Openshaw, 1995). There have been several changes to the census design and content in 2001. These are also listed in Box 4 and are discussed in detail in an edited volume by Rees *et al.* (2002).

Census cross-tabulations provide very useful information on the attributes of the population at the small area level. However, policy makers are often interested in cross-tabulations of variables that are not available from the SAS. This problem can be overcome if census microdata are available. It should be noted that population individual-level data or microdata are an invaluable resource for social science research. Compared to aggregate tabular population data, such as the SAS and LBS, population microdata contain much more detail on household or individual attributes, but are released at a coarser geographical level. The growing need for population microdata has led an increasing number of governments to commit to the decennial production of census-based microdata samples. In Britain, the release of microdata files had been under discussion since the 1970s and, after a detailed assessment of the likely risk to confidentiality, the UK census offices agreed to release Samples of Anonymised Records (SARs) from the 1991 census (Marsh and Teague, 1992; Marsh, 1993; Middleton, 1995). Similar data from the 2001 census will also be available (Dale and Teague, 2002). However, one of the major limitations of census microdata is that they are constrained by the number of questions asked in the census.

Other small area data sources

Apart from the census, there are several other sources of geographical socio-economic information. Many of these data sources are based on a combination of the census and other data. For instance, the Office for National Statistics provides the Neighbourhood Statistics online (http://neighbourhood.statistics.gov.uk/). This service offers access to a wide range of social and economic data, which include census small area data as well as data provided by local government authorities and other providers such as the Home Office (crime data) and the Land Registry (house prices). These data include information on the socio-economic characteristics of the people living in the area, as well as information on housing and crime (see Box 5). The neighbourhood statistics web site will eventually also provide 'modelled' or estimated data such as average household income at small area level. There are similar sources of information in the private sector. For example, UpMyStreet (http://www.upmystreet.com/) provides detailed socio-economic information at postcode level (see Box 6). It also provides information on local suppliers of various products and services.

Box 5 Neighbourhood Statistics (as of 16 October 2003)

- 2001 census
- Access to services
- Community well-being/social environment
- Crime and safety
- Economic deprivation
- Education, skills and training
- Health and care
- Housing
- Indices of deprivation and classifications

(Continued)

- People and society
- Physical environment
- Work deprivation

Source: http://neighbourhood.statistics.gov.uk; accessed 16 October 2003

Box 6 Private sector geographical data providers (as of 16 October 2003)

UpMyStreet presents the following information specifically about a geographical area:

- property prices
- Council Tax
- waste collections
- top schools and their exam results
- policing and crime
- local authority spending on public transport, social services and libraries
- local MP and how to contact him/her
- what's on, a local entertainment guide
- local job vacancies.

Postcode profile

ACORN profiles, a classification database devised by the market research company CACI, which businesses use to group neighbourhoods according to demographics, socio-economic profile, details of food consumption and purchasing habits.

Source: http://www.upmystreet.com/; accessed 16 October 2003

Survey data

There are numerous socio-economic surveys of households and individuals that can also be used in a geographical context. The outputs of most of these surveys are typically released at relatively coarse levels of geography (e.g. region or district level geographies). Among the most widely used socio-economic surveys is the New Earnings Survey – conducted by the Office for National Statistics (ONS), which records employment data of a relatively large sample of employees. Further, the Family Expenditure Survey (FES) has been widely used for policy analysis. The FES is a continuous survey of household expenditure and income carried out by the Office for National Statistics (sample size 10,000 households). It should be noted that the FES was combined with the National Food Survey from 2001 and is now known as the Expenditure and Food Survey. Among the survey's topics are:

- expenditure on goods and services, with considerable detail in the categories used
- income, including details about the sources of income
- possession of consumer durables and cars
- housing.

Further, the Family Resources Survey (FRS) is a continuous survey, which was launched in October 1992 by the Department of Social Security (ONS, 2000b; Dhanecha *et al*., 2003). The FRS has a relatively large sample size (around 26,000 households per year). The FRS includes:

- basic household and individual characteristics (tenure, ethnic origin, employment status, etc.)
- housing costs (rents, mortgages, Council Tax, water and sewerage charges, insurance)
- household income (including benefit receipt, unearned income, pensions, etc.)
- other costs (travel-to-work costs, childcare costs, maintenance payments, etc.)
- ownership of vehicles and consumer durables (ONS, 2000b; Dhanecha *et al*., 2003).

The General Household Survey (GHS), conducted by the Social Survey Division of the ONS (Essex Data Archive, 2000), is based on a sample of around 10,000 private households in Great Britain. Interviews are conducted with everyone aged over 16 in the household (around 18,000 adults). The GHS began in 1971 and data are available from 1973 onwards. The survey covers five 'core' subjects: population and family information, housing, employment, education and health. In addition, special topics are added from year to year. In 1998, these supplementary topics included smoking, drinking, hearing, contraception and day care. Questions related to older people were also repeated from earlier years, with the results published as a separate report, 'People aged 65 and over' (Essex Data Archive, 2000). The GHS can be used to explore the relationships between income, housing, economic activity, family composition, fertility, education, leisure activities, drinking, smoking and health. In addition to regular 'core' questions, certain subjects are covered periodically, such as:

- family and household formation
- health and related topics
- use of social services by older people and participation in sports and leisure activities (MIMAS, 2000; ONS, 2003a).

Further, the ONS conducts Omnibus, which is a multi-purpose survey that provides information on a wide range of topics including contraception, tobacco consumption, unused medicines, changes to family income, internet access, arts participation and attitudes to developing countries (Essex Data Archive, 2003; ONS, 2003b). Omnibus can be used to provide answers to questions of immediate interest, as well as to provide information on topics that do not require a full survey (Essex Data Archive, 2003; ONS, 2003b).

Another widely used survey for labour market analysis is the Labour Force Survey (LFS), which is conducted by the ONS on behalf of the Employment Department. The survey was biennial from 1973 to 1983. Since 1984, the LFS has been conducted annually, with around 60,000 sampled households (MIMAS, 2000; ONS, 2003c). The LFS has a panel design and every household is interviewed for five waves (ONS, 2001b). The LFS asks a range of questions relating to:

- household composition
- housing tenure
- ethnicity
- education and training
- employment, unemployment and job-search activities
- reasons for not wanting to work
- income
- labour mobility
- travel to work
- trade union membership
- current working conditions
- hours of work and health (sickness, accidents and health problems/ disabilities) (ONS, 2001b).

A common characteristic of all the surveys described above (with the exception of the LFS) is that they are cross-sectional: they survey a sample of households or individuals at one point in time.

One of the most comprehensive surveys in Britain is the British Household Panel Survey (BHPS), which is an annual survey of the adult population of the UK, drawn from a representative sample of over 5,000 households. The aim of the survey is to deepen the understanding of social and economic change at the individual and household level in Britain, as well as to identify, model and forecast such changes, their causes and consequences in relation to a range of socio-economic variables (Taylor *et al*., 2001). Boxes 7 and 8 outline the core household and individual questions asked in the BHPS questionnaires. These questions have generated a wealth of socio-economic and demographic variables, which make the BHPS unique in that it includes almost all the variables contained in most other national social survey data in Britain.

Box 7 BHPS Individual Questionnaire: details of the core, rotating core and variable component question subject areas (Taylor *et al.*, 2001)

Core

1 Neighbourhood and individual
- Demographics
- Birthplace, residence
- Satisfaction with home/neighbourhood
- Reasons for moving
- Ethnicity
- Educational background and attainments
- Recent education/training
- Partisan support
- Changes in marital status
- Citizenship

2 Current employment
- Employment status
- Not working/seeking work
- Self-employed
- Sector private/public
- SIC/SOC/ISCO
- Nature of business/duties
- Workplace/size of firm
- Travelling time
- Means of travel
- Length of tenure
- Hours worked/overtime
- Union membership
- Prospects/training/ambitions
- Superannuation/pensions
- Attitudes to work/incentives
- Wages/salary/deductions
- Childcare provisions

(Continued)

- Job-search activity
- Career opportunities
- Bonuses
- Performance-related pay

3 Finances
- Incomes from: benefits/allowances/pensions/rents/savings/interest/dividends
- Pension plans
- Savings and investments
- Material well-being
- Consumer confidence
- Internal transfers
- External transfers
- Personal spending
- Roles of partners/spouses
- Domestic work/childcare/bills/everyday spending
- Car ownership/use/value of car
- Interview characteristics
- Windfalls

Rotating core

1 Health and caring
- Personal health condition
- Employment constraints
- Visits to doctor
- Hospital/clinic use
- Use of health/welfare services
- Social services
- Specialists
- Check-ups/tests/screening
- Smoking
- Caring for relatives/others
- Time spent caring for others
- Private medical insurance
- Activities in daily living

(Continued)

2 Employment history
- Past year
- Labour force status spells
- Size / sector / nature of business / duties
- Wages / salary / deductions
- Reasons for leaving / taking jobs

3 Values and opinions
- Partisanship / interest in politics
- Religious involvement
- Parental questionnaire

Variable components

1 Lifetime marital status history (wave 2)
- Number of marriages
- Marriage dates
- Divorce / widowhood /
- Separation dates
- Cohabitation before marriage

2 Lifetime employment status history (wave 3)
- Start and finish dates
- Labour force status
- Sector / nature of business duties

3 Health and caring
- Children's health
- Other health scales: SF36 (wave 9)

4 Computers and computing (wave 6/7)
- Ownership and usage

5 Lifetime fertility and adoption history (wave 2 and wave 8 catch-up)
- Birth dates
- Adoption dates
- Gender of children
- Leaving or mortality dates

(Continued)

6 Lifetime cohabitation history (wave 2 and wave 8 catch-up)
- Start and finish dates
- Number of partners

7 Neighbourhood and demographics
- Driving licence
- Parents' employment background
- Family background
- Difficulties with debt
- Community and neighbourhood

8 Employment (wave 9)
- National Minimum Wage
- Work strain
- Work orientation

9 Lifetime employment status history (wave 2)
- Start and finish dates
- Employment status

10 Values and opinions
- Aspirations for children
- Important events
- Quality of life

11 Credit and debt
- Investment and savings
- Commitments

12 Crime
- Criminal activity on local area
- Perceptions of crime

Box 8 BHPS Household Questionnaire: details of the core question subject areas (Taylor *et al.*, 2001)

1 Size and condition of dwelling

2 Ownership status

3 Length of tenure

4 Previous ownership

5 Interview characteristics

6 Household finances
- Rent and mortgage, loan and HP details
- Local authority service charges
- Allowances/rebates
- Difficulties with rent/mortgage payments

(Continued)

- Household composition
- Consumer durables, cars, telephones, food
- Heating/fuel types, costs, payment methods
- Non-monetary poverty indicators

However, as is the case with all other large surveys, BHPS gives information at relatively coarse levels of geography.

Table 2 shows the geographical distribution of the households in the BHPS wave 1.

The remainder of this book describes a method that adds a geographical dimension to survey microdata by combining them with small area census data.

Table 2 Origin of wave 1 BHPS households (AREGION)

Value label	Frequency	Frequency (%)
Inner London 1	498	5.8
Outer London 2	597	7
Rest of South East 3	1,611	18.9
South West 4	713	8.4
East Anglia 5	303	3.6
East Midlands 6	595	7
West Midlands Conurbation 7	391	4.6
Rest of West Midlands 8	369	4.3
Greater Manchester 9	396	4.6
Merseyside 10	195	2.3
Rest of North West 11	363	4.3
South Yorkshire 12	197	2.3
West Yorkshire 13	299	3.5
Rest of Yorks and Humber 14	257	3
Tyne and Wear 15	202	2.4
Rest of North 16	293	3.4
Wales 17	392	4.6
Scotland 18	853	10

Part III
Practical application

9 How to add geography to national survey data

The datasets described in the previous chapter can be very useful for policy analysis and research. For instance, census data can be used to map the geographical distribution of socio-economic groups and to monitor past trends (e.g. using data from the 1971, 1981 1991 and 2001 censuses). Neighbourhood statistics can be used in the same way. On the other hand, survey data such as the Labour Force Survey or the BHPS can be used to paint more detailed pictures of the life of households in Britain. However, it is possible to add greater value to all these datasets by combining them in order to create powerful geographical databases that can be used for policy analysis at various geographical scales. This chapter shows how geographical detail can be added to survey data.

Reweighting survey household records

All the households in the BHPS are given a weight that compensates for error, bias, refusals, etc. These weights can be readjusted in order to fit small area descriptions, such as census small area data (the weights can be readjusted so that they would add up to these small area descriptions). An example of how such a readjustment can be carried out is described in Tables 3–6. In particular, Table 3 gives a hypothetical individual microdataset comprising five individuals, who fall within two age categories. Further, Table 4 depicts a small area statistics table for a hypothetical area, whereas Table 5 depicts a cross-tabulation of the hypothetical microdataset, so that it can be comparable to Table 4.

Table 3 A hypothetical microdataset (original weights: Table *w*)

Individual	Gender	Age group	Weight
1st	Male	Over-50	1
2nd	Male	Over-50	1
3rd	Male	Under-50	1
4th	Female	Over-50	1
5th	Female	Under-50	1

Table 4 Hypothetical small area data tabulation (Table *s*)

Age	Male	Female
Under-50	3	5
Over-50	3	1

Table 5 The hypothetical microdataset, cross-tabulated by age and sex (Table *m*)

Age	Male	Female
Under-50	1	1
Over-50	2	1

Using these data, it is possible to readjust the weights of the hypothetical individuals, so that their sum would add up to the totals given in Table 3. In particular, the weights can be readjusted by multiplying them by the value of the cell in Table 4, which denotes the category in which they belong over the respective cell in Table 5. This can be expressed as follows:

$$n_i = w_i \times s_{ij} / m_{ij}$$

where n_i is the new household weight for household i, w_i is the original weight for household i, s_{ij} is element ij of Table s (small area statistics table, which is the equivalent of Table 4) and m_{ij} is element ij of Table m (reproduced table using the household microdata original weights, which is the equivalent of Table 5 in the example).

Table 6 depicts how this simple formula is used to readjust the weights of the individuals in the example.

The above process can then be used to reweight the individuals to fit another table. In the context of this book, this reweighting procedure was adopted iteratively to readjust the BHPS households weights[1] so that they would fit the small area statistics tables described in the previous chapter. The generated weights for each household represent the probabilities of BHPS households to 'live' in a given ward.

After generating the BHPS household weights for each small area, the next step is to select the appropriate households (or, in other words, convert the decimal weights or probabilities into integer weights). In the context of the research presented here, different 'integer weighting' or integerisation methodologies were adopted and tested, and it was concluded that the following methodology represented the best solution.

Define two variables named *counter* and *weight*, set them to zero and then do the following.

- Sort all households into ascending order of probability of living in the small area (which were calculated using the method described above) being populated.
- Increase cumulative *weight* by the *weight* (probability) of the next sorted household *h*(*counter*). For instance, if *counter* = 0, the *weight* is increased by the probability of the first household: *h*(0).
- If cumulative *weight* > 1 give to the household *h*(*counter*) an integer weight equal to the rounded *weight* value and subtract this value from *weight* (e.g. if *weight* = 2.05 set household *weight* = 2 and set *weight* = 2.05 – 2 = 0.05). Increase *counter* by 1 (move to next household).

Table 6 Reweighting the hypothetical microdata set in order to fit Table 4

Individual	Gender	Age group	Weight	New weight
1st	Male	Over-50	1	$1 \times 3/2 = 1.5$
2nd	Male	Over-50	1	$1 \times 3/2 = 1.5$
3rd	Male	Under-50	1	$1 \times 3/1 = 3$
4th	Female	Over-50	1	$1 \times 1/1 = 1$
5th	Female	Under-50	1	$1 \times 5/1 = 5$

- If *counter < total number of households in the small area*, return to 2, else exit.

The implementation of the above simple algorithm creates a small area level microdataset. However, the method described above will give the appropriate weights for the households to be representative of the small area on the basis of the census tables used as 'constraints' in the exercise. These weights may translate into relatively high overestimates and underestimates of some variables that were not used as constraints in the simulation. In order to tackle this problem, an algorithm aimed at swapping suitable simulated households between wards in order to further reduce the error was developed. The steps taken to reduce the error were as follows.

- Identify wards with the highest overestimate and underestimates for each variable.
- Compare each household in the simulated database with all other households and search for households that have all attributes in common but one.
- For each pair of almost identical households swap the households between the areas with the highest overestimate and underestimate.
- Move to the next household most over- or under-represented and repeat the process.

The implementation of this algorithm led to the reduction of error for unconstrained variables in some areas.

It should be noted that the generated geographical microdata can be used to draw a picture of how different types of households live in different localities. In particular, it is possible to use the model outputs to generate small area descriptions of typical households such as the one shown in Table 7.

In effect, 'real people' are imputed into census areas. This way of presenting the simulation outputs can provide useful insights into social policy design, as it may enable policy makers to identify what kind of households will be affected by different types of policies. This is further discussed in Chapter 14.

Table 7 Example of small area descriptions of typical households generated by the model outputs

Age of household head(s)	Description
72	Single-person household, female, retired assembler/lineworker (electrical/electronic goods). Feels that is *just about getting by* financially. House owned outright. Believes that all health care should be free.
56 and 52	Married couple, male aged 56, economically active but unemployed. Formerly employed as motor mechanic/auto engineer. Female aged 52, economically inactive. Food expenditure £25 per week. No car. House owned with mortgage. Highest educational qualification of male: GCE O levels. Female has no formal qualifications.

(Continued overleaf)

Table 7 Example of small area descriptions of typical households generated by the model outputs (continued)

Age of household head(s)	Description
52 and 49	Married couple, one child (aged 7). Household has two cars. Both parents fully employed on permanent posts. Male has university higher degree and an associate professional occupation. Female has university degree and works as a teacher. House owned with mortgage. Weekly food expenditure £50.
26	Single-person household. Higher qualification, one car, works in catering. Privately rented furnished accommodation.
31 and 31	Married couple, three children (7, 5 and 2). Male has higher qualification and permanent employment as a sales representative. Female is currently economically inactive, but has A-level qualifications and working experience as a receptionist. The household has two cars. Weekly food expenditure £70. Tenure status: house owned with a mortgage.
29 and 25	Married couple with two children aged 5 and 1. Male has higher qualification and female has O levels. Male is employed full time as a coach and vehicle body builder and female is employed full time as a financial clerk. Household has two cars. Tenure status: house owned with a mortgage. Household weekly food expenditure: £40.
35 and 30	Married couple with two children aged 7 and 4. Both parents in full employment. Male works as a production manager in manufacturing, construction and energy industry and his highest formal qualification is A levels. Female has O levels and works full time as a plant/machine operative. Weekly food expenditure: £50. Household has one car. House owned with a mortgage.
19 and 17	Cohabiting couple. Male's highest qualification is CSE grade 2–5. He is employed on a full-time basis on a temporary post (associate professional and technical occupation). Female's highest qualification is A levels. She is employed full time on a permanent job in childcare and related occupations. Weekly food expenditure: £35. Tenure: privately rented furnished. One car.
39 and 32	Married couple, three children (aged 11, 9 and 0). One car. Male employed full time on a permanent post as an assembler/lineworker. Female economically inactive (family care). Weekly food expenditure: £20.
44 and 42	Married couple with a 21-year-old son living with them. Three cars. All household members in full employment. Male employed in food trade, female employed as a clerk, 21-year-old son employed as a plant and machine operative. House owned with a mortgage.
34	Separated one-person household. Occupied as a manager and proprietor in service industry. Has one car and lives in privately rented accommodation. Weekly food expenditure: £75.
50 and 40	Married couple with no children. Male works as hotel and accommodation manager, female works as a buyer and purchasing officer. Tenure status: house owned outright. Household has one car and spends £50 weekly on food.

(Continued)

Table 7 Example of small area descriptions of typical households generated by the model outputs (continued)

Age of household head(s)	Description
40 and 34	Food £45, male has university degree, female has apprenticeship. Female works as an assistant nurse, male works as a mechanical engineer. Two cars. Tenure: owned with mortgage.
32 and 28	Married couple. Male has postgraduate university degree, female has university first degree. They have two cars and spend £30 weekly on food. Both work as university teaching professionals. Tenure: house owned with a mortgage.
34 and 32	Married couple with two children (aged 12 and 10). £40 food. Male has no formal qualifications and works as a plant and machine operator. Female has a university degree and has a professional occupation. The household has two cars. Tenure status: house owned with mortgage.
35 and 36	Married couple with no children. Female is 36 years old and has a university degree and a professional full-time job. Male is 35 years old and works in sales occupations (highest educational qualification: A levels). The household has two cars and spends £35 weekly on food. Tenure status: house owned with a mortgage.

10 Updating small area population statistics

The previous chapter described how it is possible to add geography to a survey dataset on the basis of small area data at one point in time. It is also possible to make the method described in the previous chapter more dynamic by applying it on the basis of projected small area data.

There are several methods of projecting small area population data into the future. This chapter shows how it is possible to project small area census data into future years on the basis of past trends. In order to do so, digital data from three UK censuses were used. In particular, projections of small area statistics tables were calculated using the 1971, 1981 and 1991 census Small Area Statistics (SAS). Using these three time points, a trend curve was produced allowing tables to be predicted up to 2021. The projections of small area statistics tables were undertaken at ward level. However, to avoid problems of wards with low populations, the ward data were smoothed. For the table for each ward, the change that has occurred for the 20,000 population nearest to that ward was used. This change was then applied to the values for each ward.

Each table has three categories of households. Projections were calculated using proportions of households in each category in each ward. The first category was calculated as in the equations below, the second category was then calculated as the proportion that category takes up of what is left, using the following equations:[1]

Projections for 2001

$$A = \exp(\text{In } W \times (\text{In } w)^2 \times \text{In } u / (\text{In } v))^3$$

Projections for 2011

$$B = \exp(\text{In } A \times (\text{In } x)^2 \times \text{In } v / (\text{In } w))^3$$

Projections for 2021

$$C = \exp(\text{In } B \times (\text{In } y)^2 \times \text{In } w / (\text{In } x))^3$$

where:

u = smoothed proportion in 1971
v = smoothed proportion in 1981
w = smoothed proportion in 1991
x = smoothed proportion in 2001
y = smoothed proportion in 2011
W = ward proportion in 1991
A = ward proportion in 2001
B = ward proportion in 2011
C = ward proportion in 2021.

The proportion in the third category was then calculated as 1 minus the proportions in the two other categories to ensure that all households were allocated to a category.

Once projections of the proportions of households that fall into each table category had been calculated, they needed to be applied to ward-level household projections to produce estimates of the numbers of households in each ward in each category. There are no official household projections at ward level; therefore a method had to be devised. Projections were required every ten years between 1991 and 2021. The DTLR publishes projections of total households for counties in England for every five years up to 2021. These county totals can be used to produce ward-level estimates of households. It was assumed that the distribution of households between wards in each county would remain the same as in 1991. This 1991 distribution can be calculated from the 1991 census of population. For each county, the proportion of its households in each ward is calculated. This proportion is assumed to remain stable up to 2021. Therefore ward-level household projections can be produced by multiplying the DTLR household projections for counties by the proportion of households for each ward. It should be noted that these projections would provide a baseline population. It may also be possible to provide alternative projections on the basis of local knowledge when this is available.

For the purposes of this book, the data were projected constrained by six small area tables, each with three categories. The tables and their categories are listed in Table 8.

It should be noted that the class composition table is actually a subset of the employment table, i.e. class is allocated only to households with an economically active head. The three class categories are made up from 1991 Socio-economic Groups (SEGs): the affluent group comprises SEGs 1, 2, 3, 4 and 13; the middle group is SEGs 5, 8, 9, 12, 14, 16 and 17; and the poor group is made up of SEGs 6, 7, 10, 11 and 15.

After applying the methodology described above, it is possible to generate some illustration of projections of socio-economic groupings in Britain. Figures 2–10 show the projected trends in the geographical distribution of these groups.

As can be seen in Figures 2–6, there is an estimated increasing spatial polarisation of the households classified as *poor*. In particular, as seen in Figure 2, most of the areas with relatively high concentrations of *poor households* are clustered in the North of Britain and the areas with the lowest concentrations of *poor households* are in the South East. Nevertheless, it is worth noting that, as can be seen in the Figure 2 insets, there are several areas in 1991 with relatively high concentrations of poor households in east and inner city London. There are also relatively high concentrations of this group of households in the North of England and, in particular, in inner-city Leeds, Sheffield, Manchester and Liverpool. However, it is interesting to note that there is a projected change in these patterns in 2001, 2011 and 2021. In particular, there is a significant decrease in the proportions of *poor households* in London and, in contrast, increases of *poor households* in the North of England. Overall, there is a significant increase in the proportions of *poor households* in the North of England, and in most areas of Scotland and North Wales. Figure 6 summarises the projected change in the geographical distribution of poor households between 1991 and 2021.

In contrast to the distributions of *poor households*, the areas with high concentrations of

Table 8 Constraint tables

Table	Categories		
Car ownership	No cars	1 car	2+ cars
Class composition	Affluent	Middle-class	Less affluent
Demography	1 child	2+ children	No children
Employment	Economically active	Retired	Inactive
Household composition	Married couple	Lone parent	Other
Tenure	Owner-occupied	Council tenants	Other

Figure 2 Spatial distribution of poor households in 1991

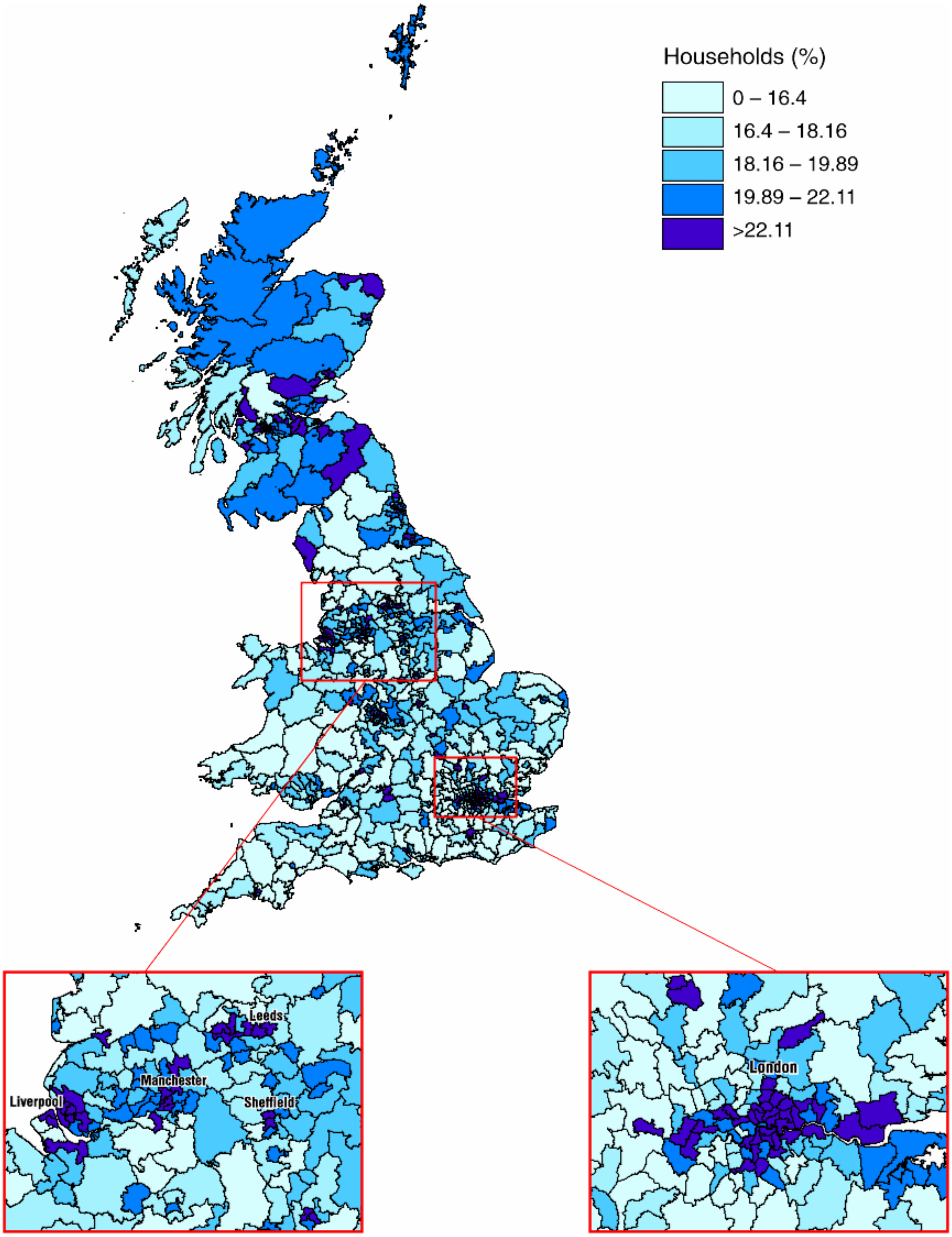


Figure 3 Projected spatial distribution of poor households in 2001

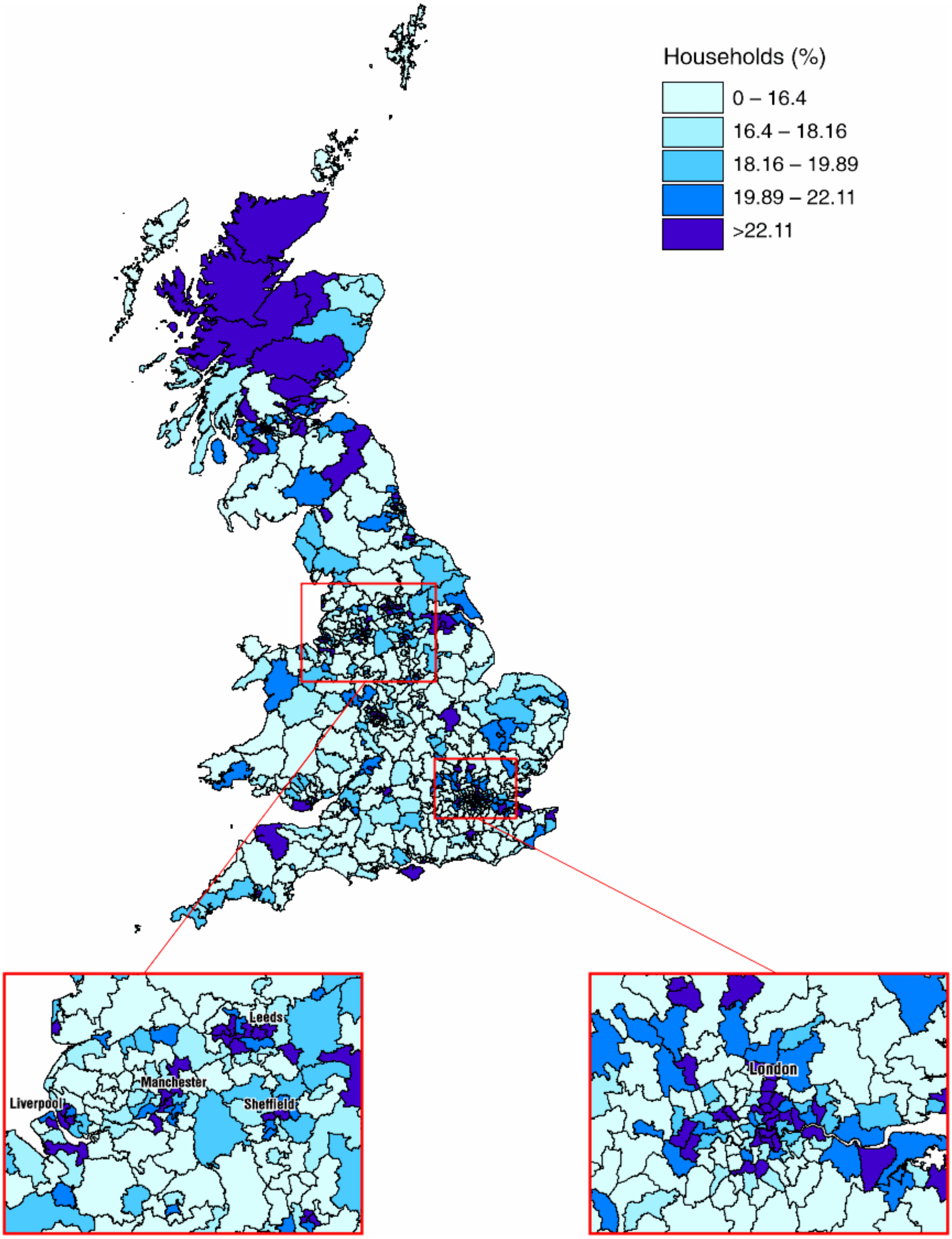


Figure 4 Projected spatial distribution of poor households in 2011

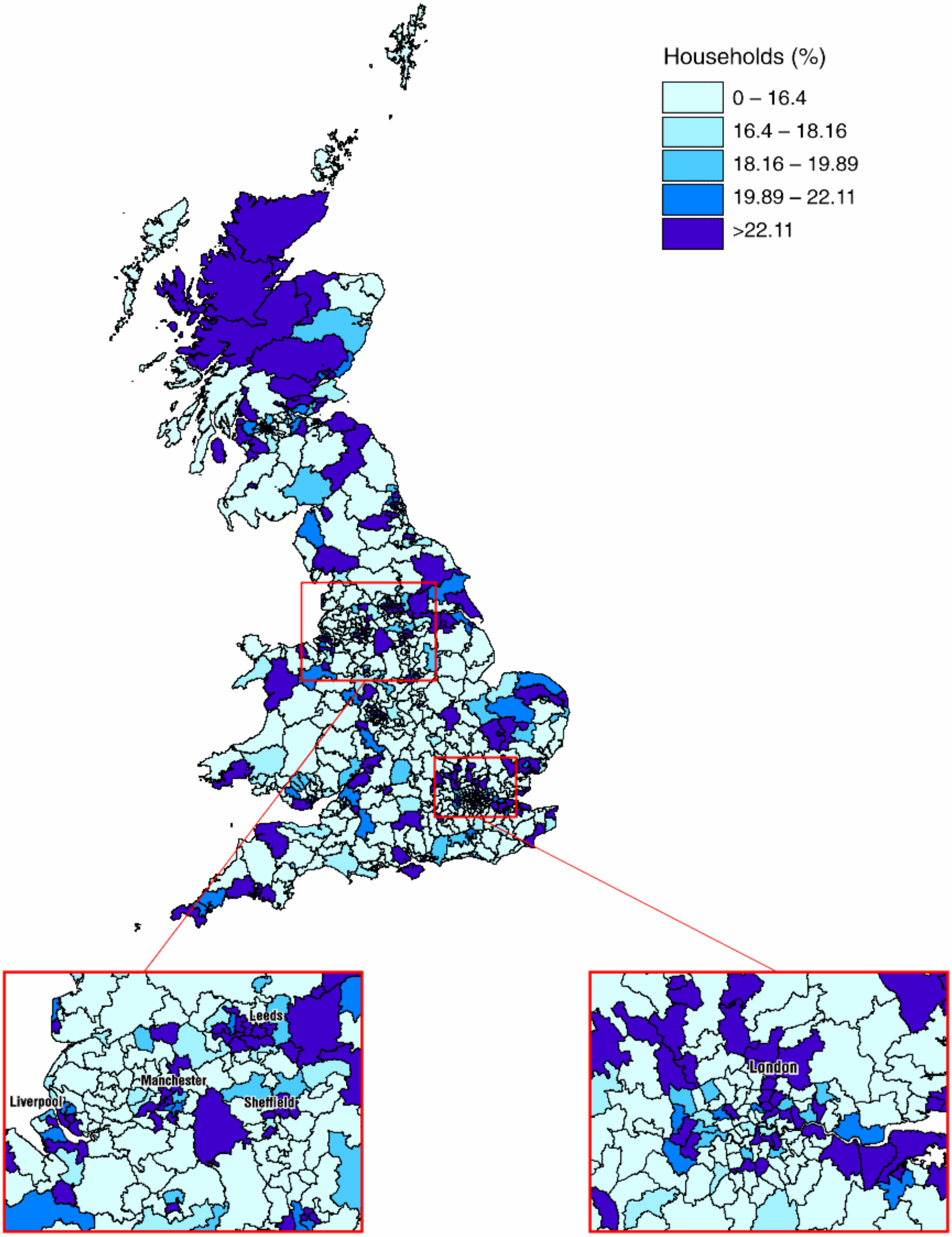


Figure 5 Projected spatial distribution of poor households in 2021

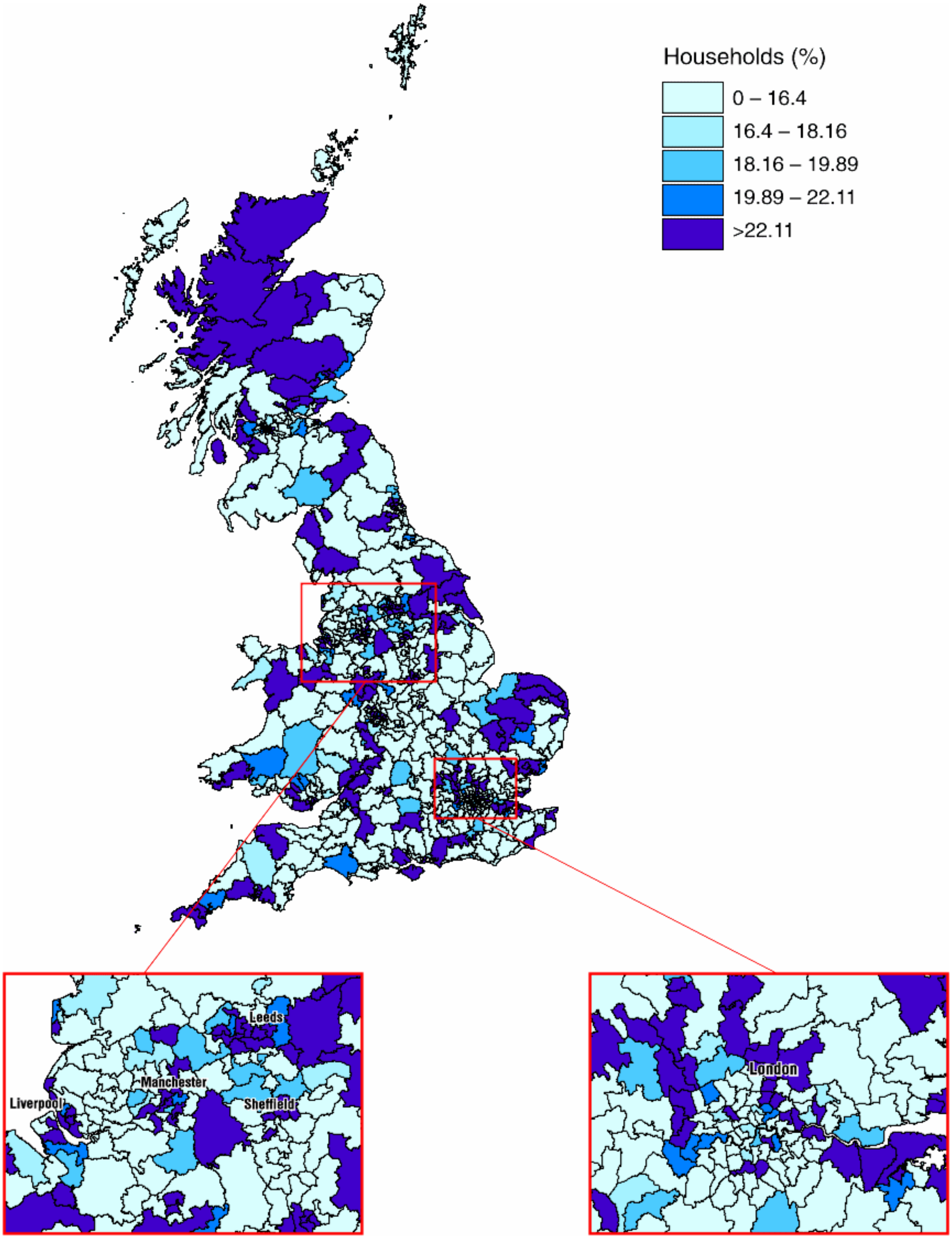


Figure 6 Changes in the distribution of poor households, 1991–2021 (% of households in 2021 minus % of households in 1991)

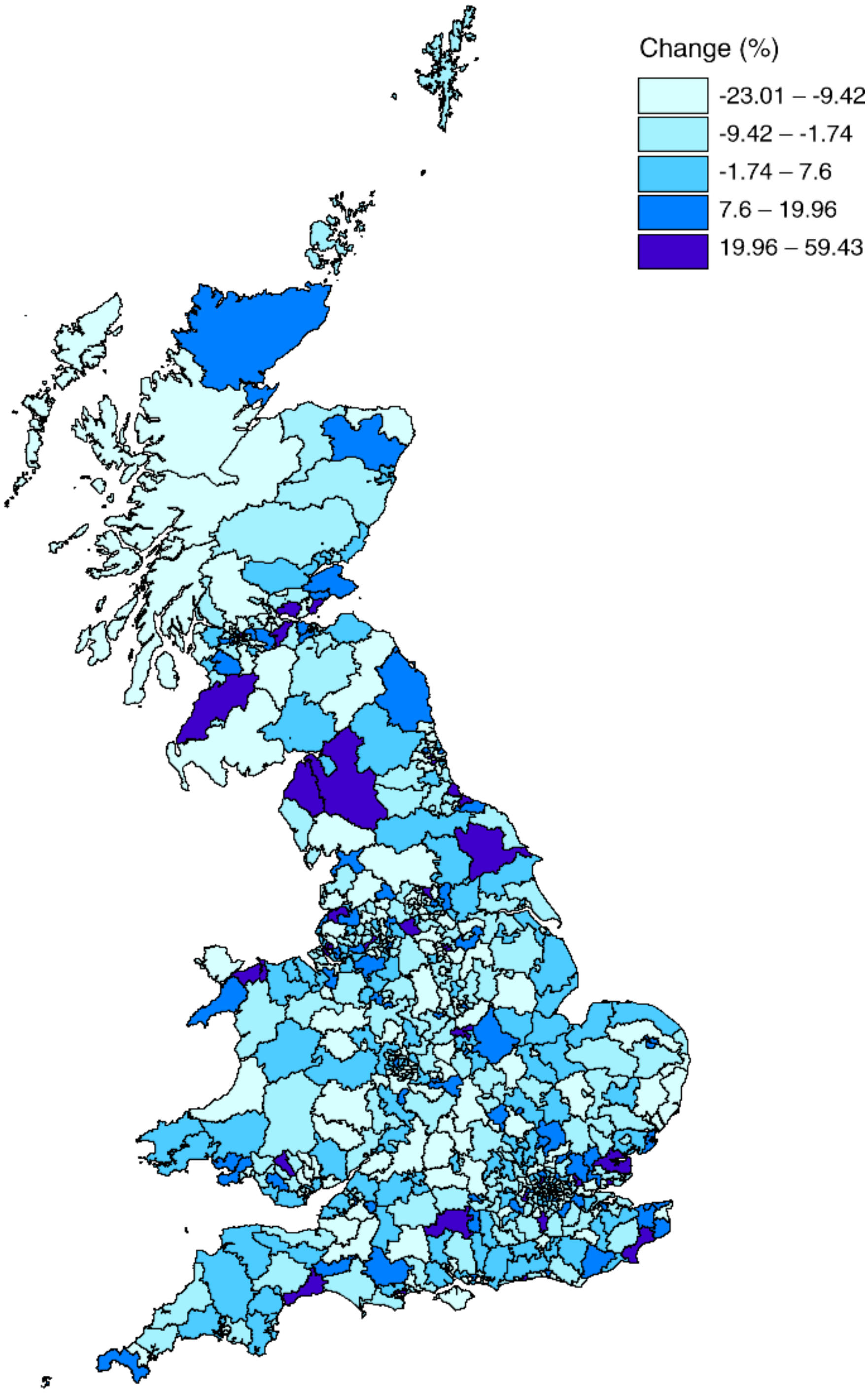


Figure 7 Spatial distribution of affluent households in 1991

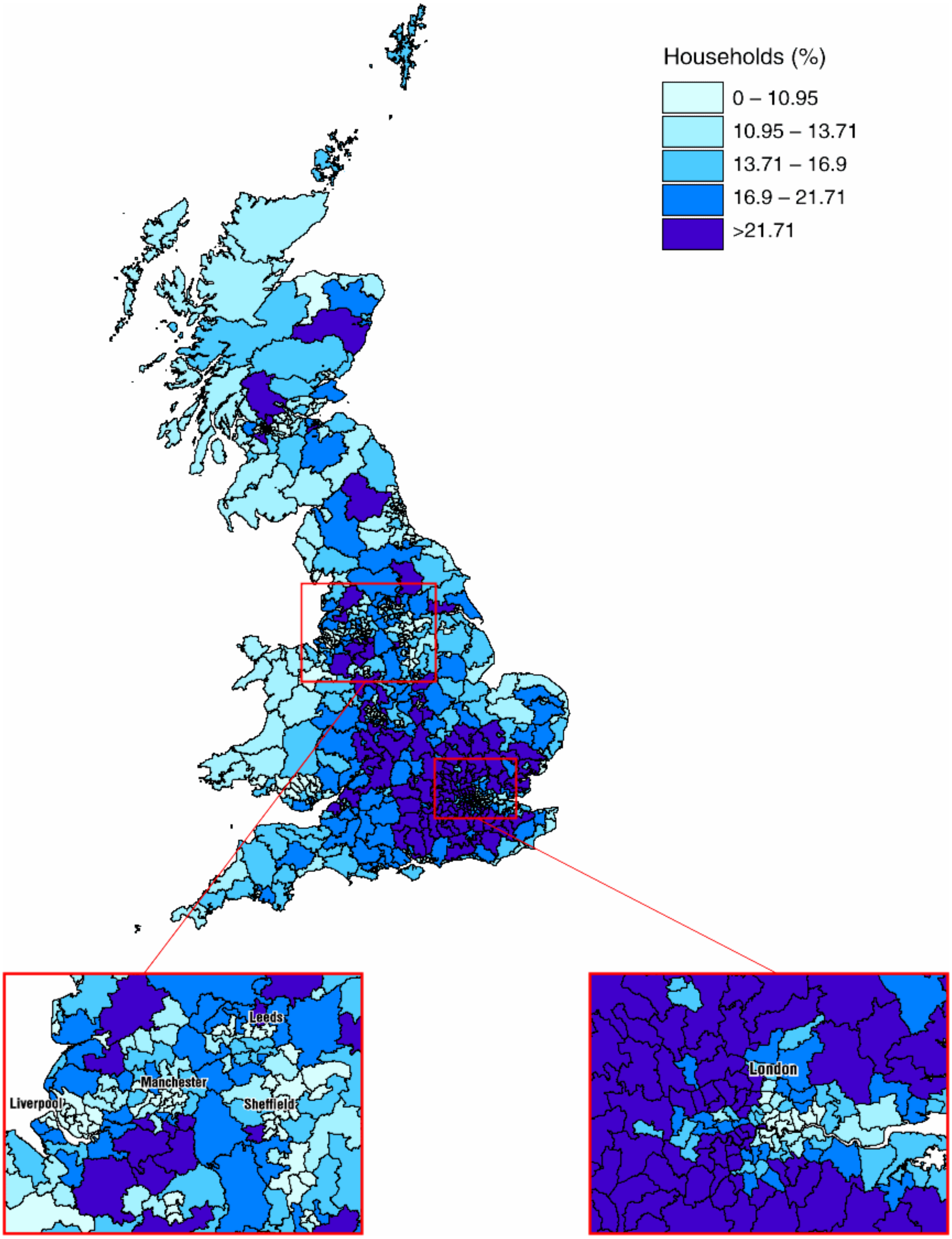


Figure 8 Spatial distribution of affluent households in 2001

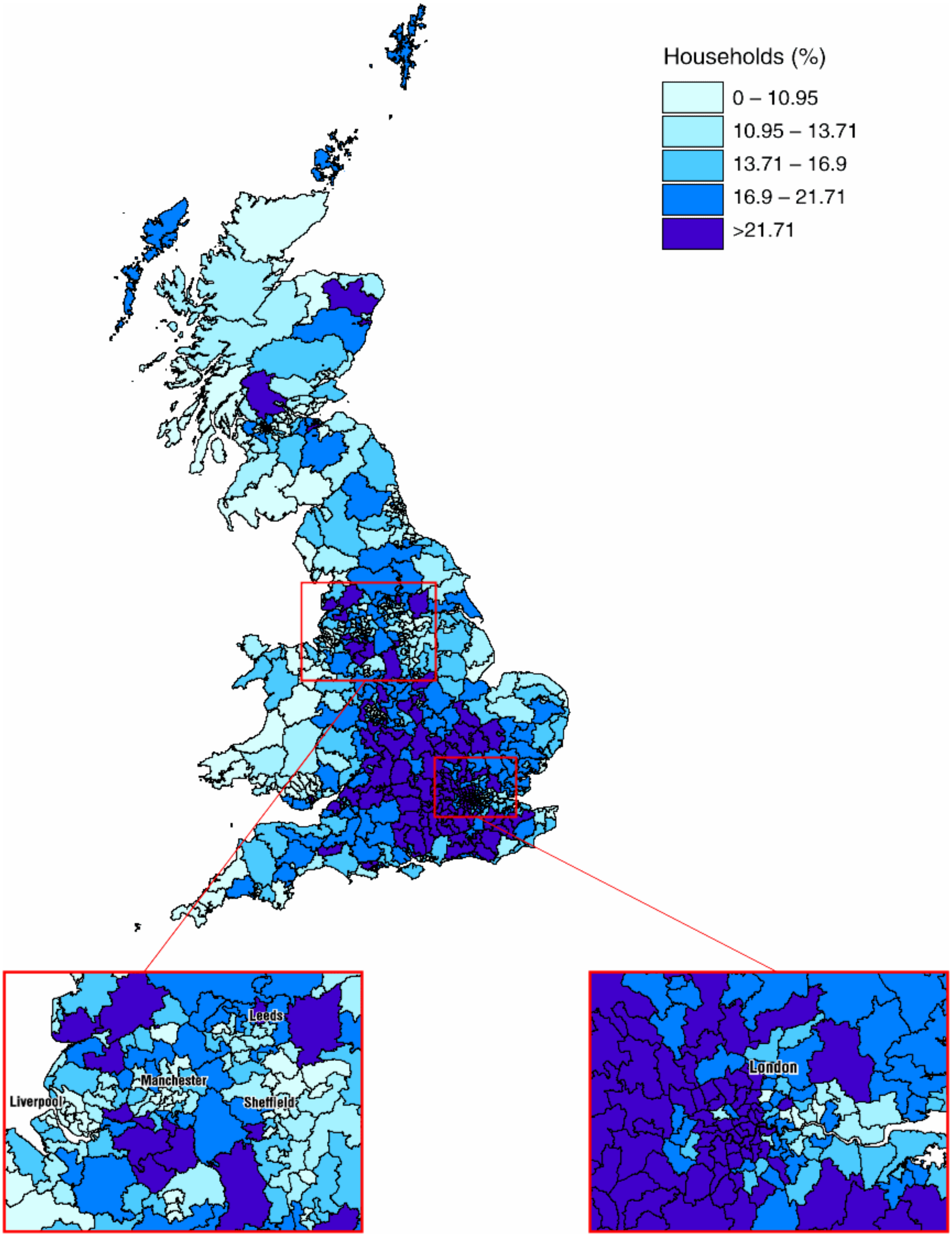


Figure 9 Spatial distribution of affluent households in 2011

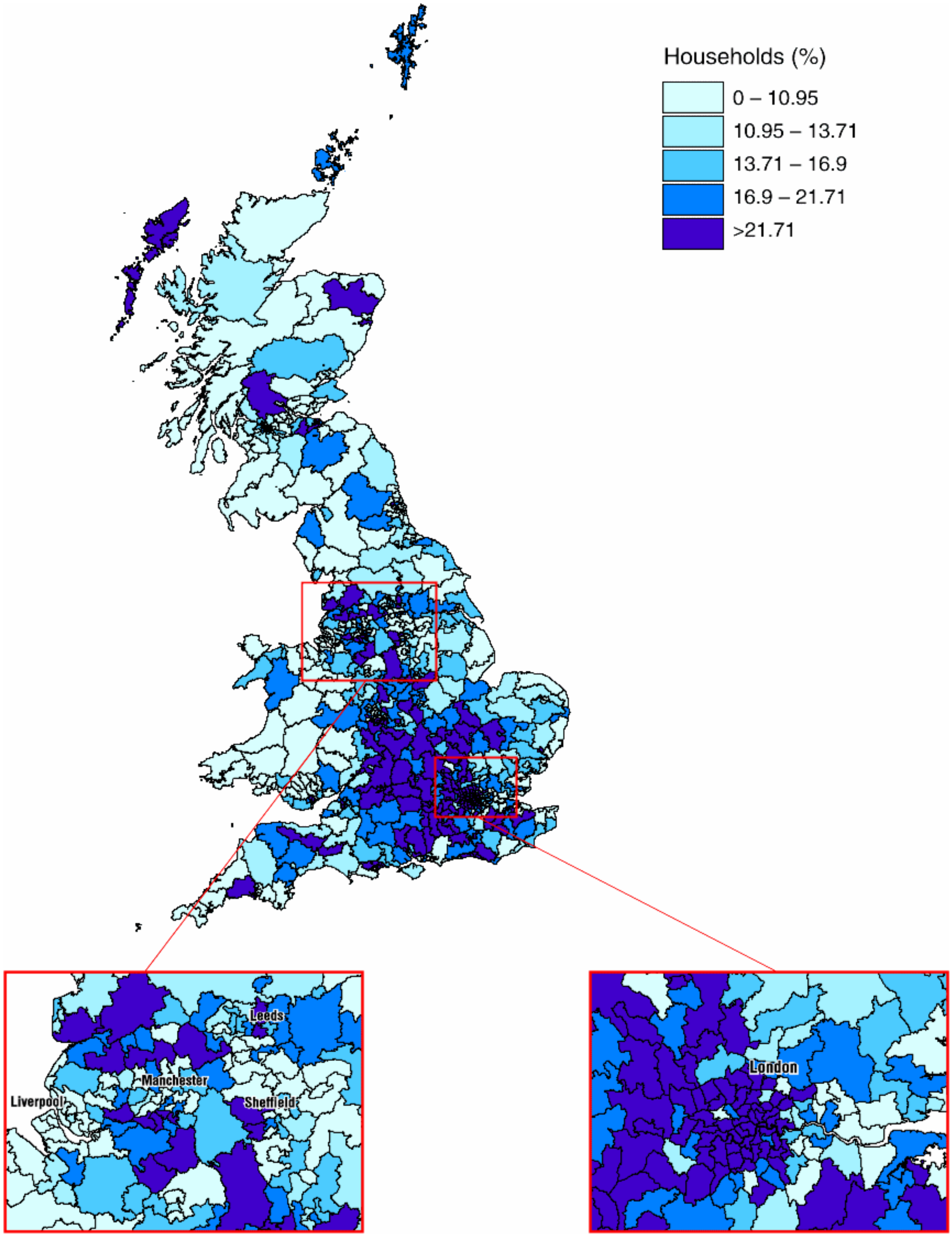
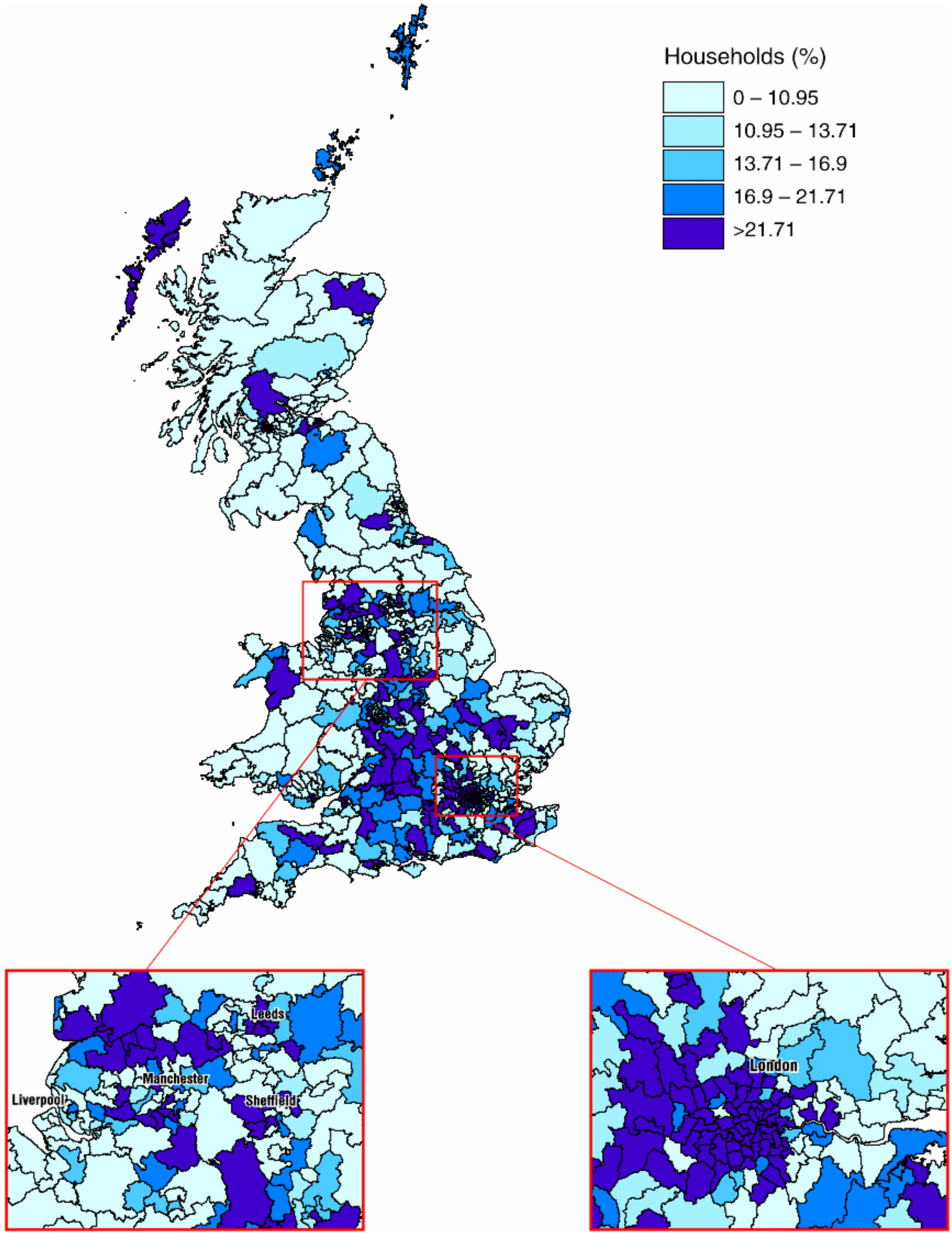


Figure 10 Spatial distribution of affluent households in 2021



affluent households are mainly clustered in the South and South East. There are also a few areas with high proportions of *affluent households* in north-west Scotland and in some parts of the Midlands and the North of England. Further, the projection results suggest an estimated polarisation between 2001 and 2021 with *affluent households* increasingly concentrating in the South East and in the Midlands. The London inset in Figures 7–11 suggests a polarisation of affluent groups in this area, as there are significant increases in the proportions of *affluent households* in central London as well as in the west, south and north of the city. In contrast, there are decreases in the proportions of affluent households in the east parts of the city. Similar polarisation trends are projected for the North of England, as can be seen in the second inset in Figures 7–10. In particular, the proportions of *affluent households* are projected to increase significantly in some suburban areas and generally decrease in all other areas. Figure 11 summarises the projected change in the geographical distribution of affluent households between 1991 and 2021.

Figures 12–15 show the actual spatial distribution of households with a *retired* head in 1991 and the projected distributions in 2001, 2011 and 2021. According to these projections, *retired households* will move out of London and big cities, and are projected to concentrate in more rural as well as coastal areas in Britain.

Figure 11 Changes in the distribution of affluent households, 1991–2021 (% of households in 2021 minus % of households in 1991)

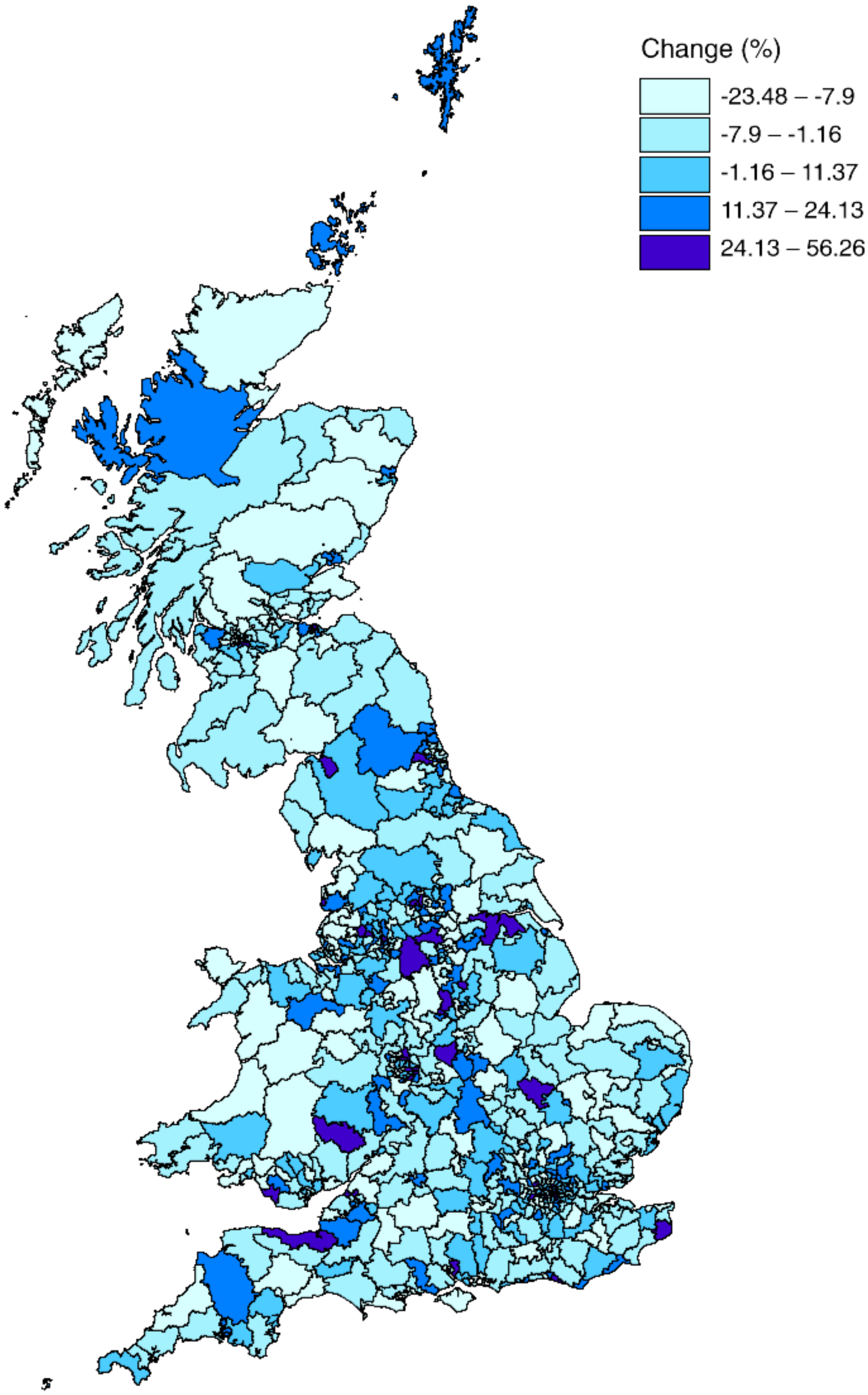


Figure 12 Spatial distribution of households with a retired head in 1991

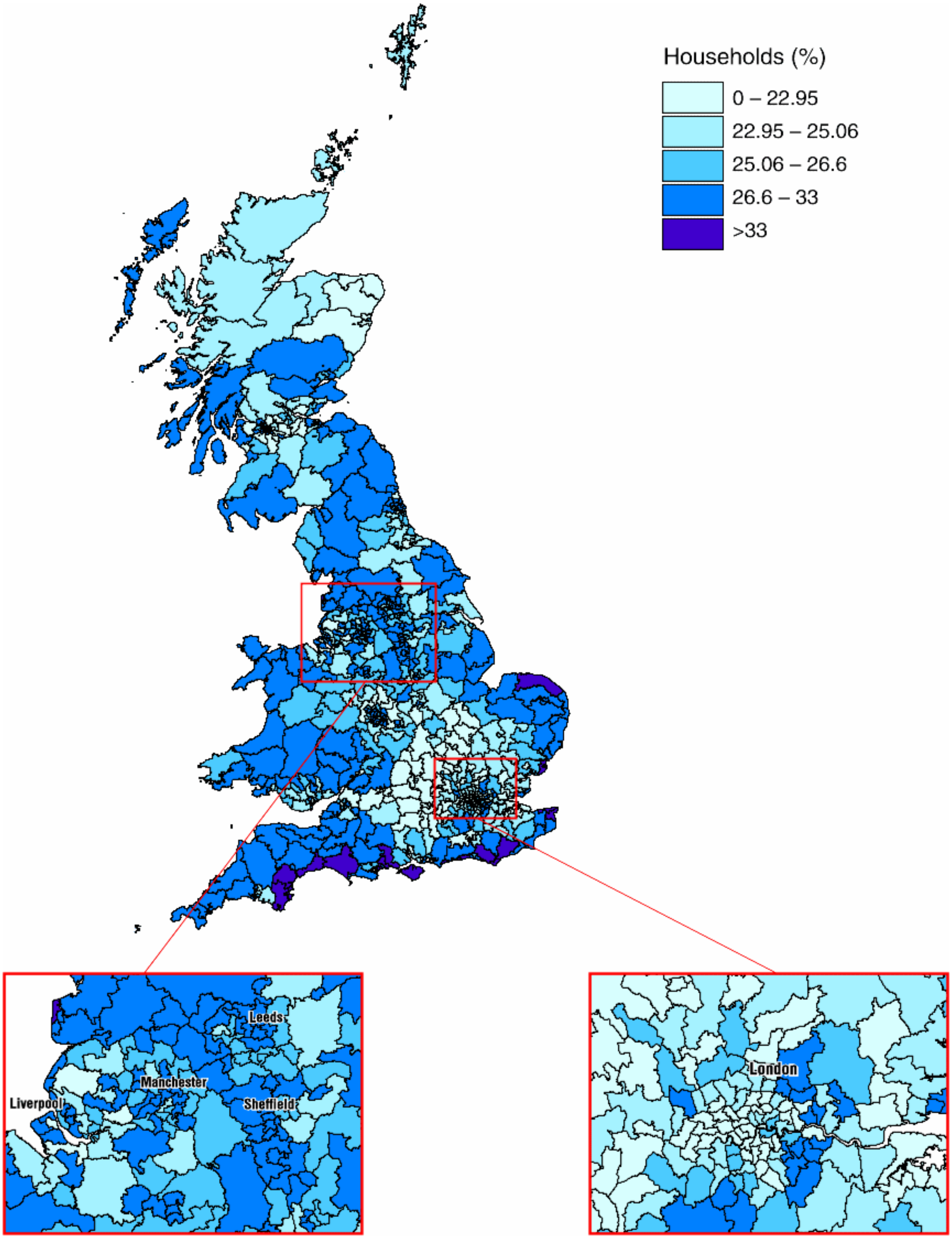


Figure 13 Projected spatial distribution of households with a retired head in 2001

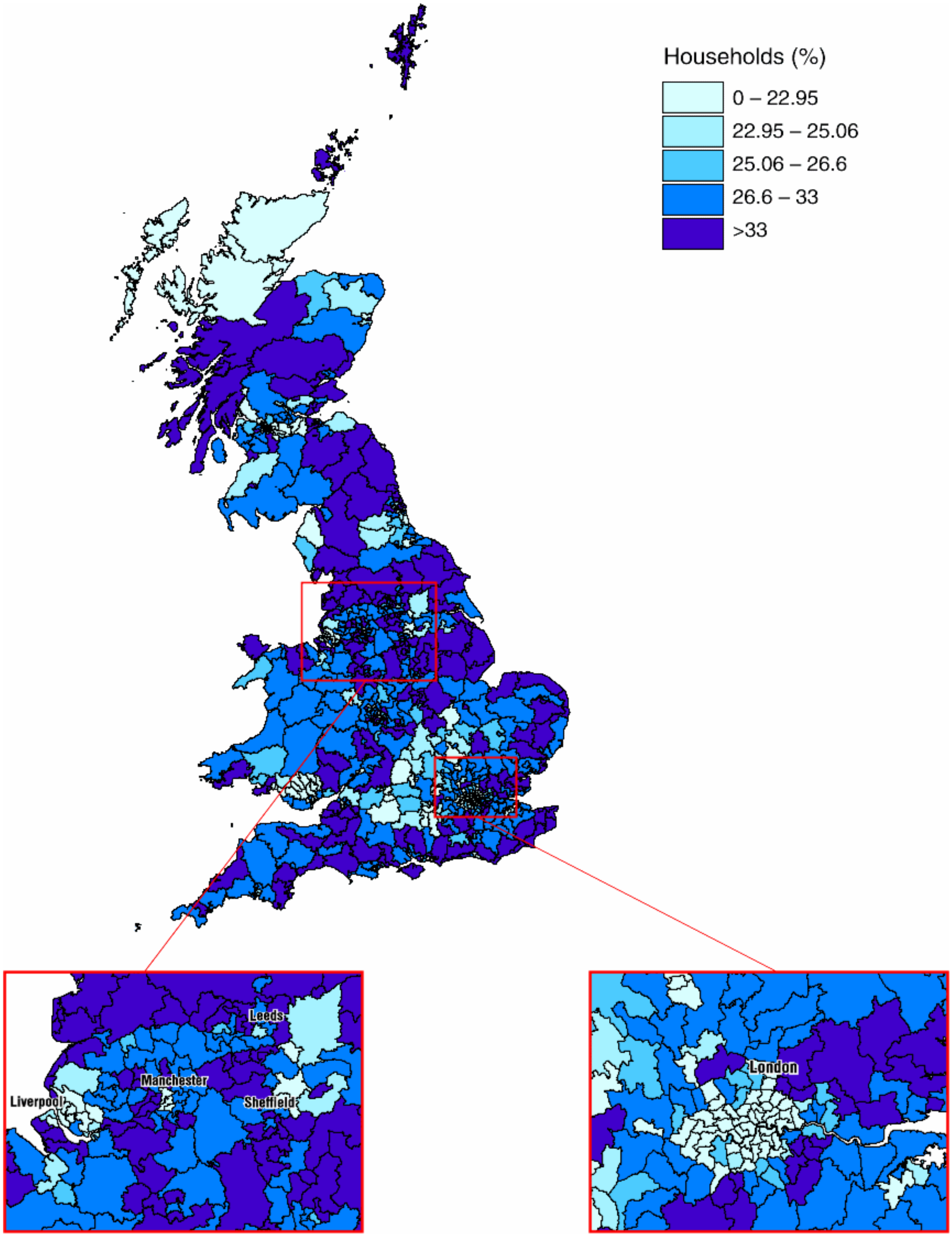


Figure 14 Projected spatial distribution of households with a retired head in 2011

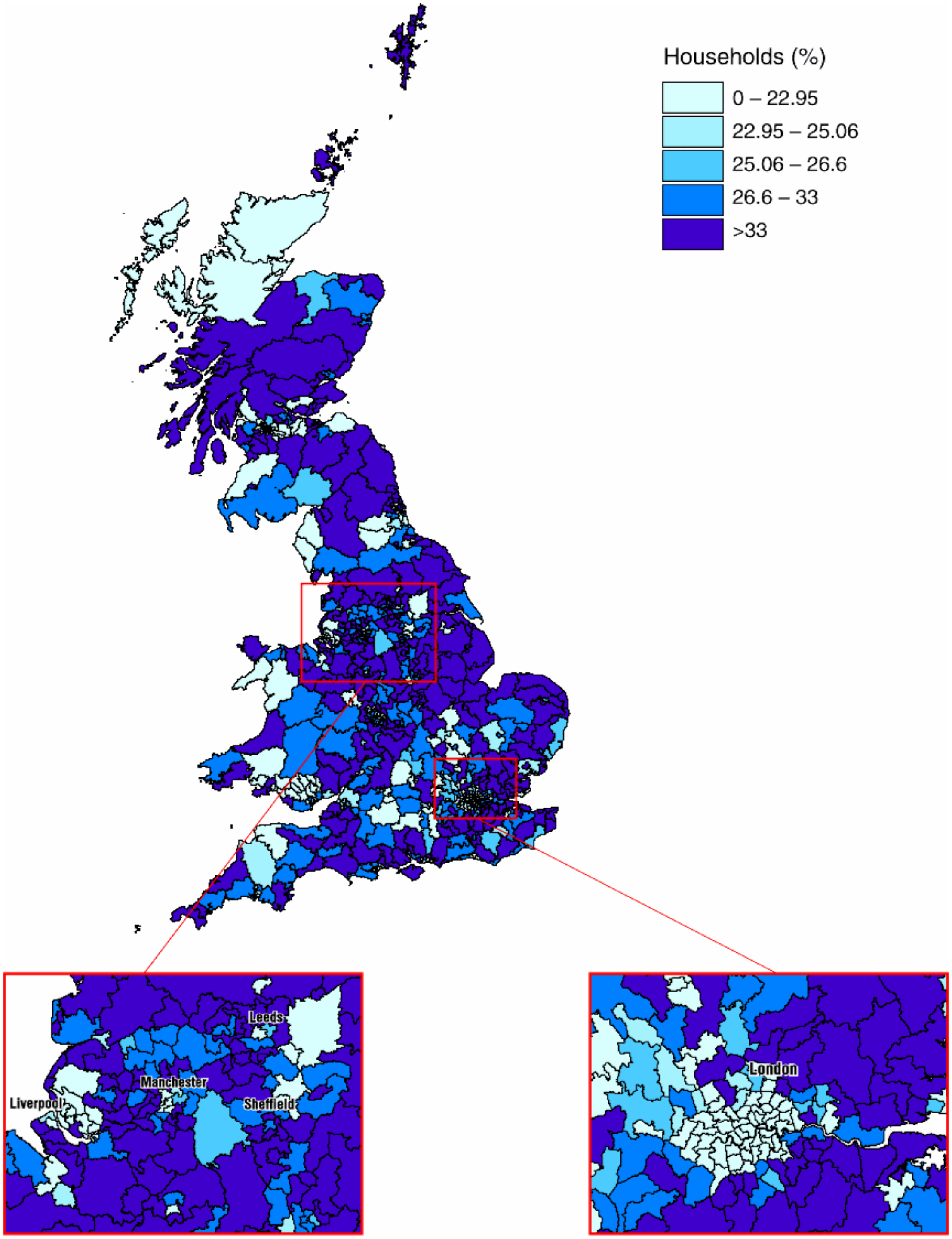
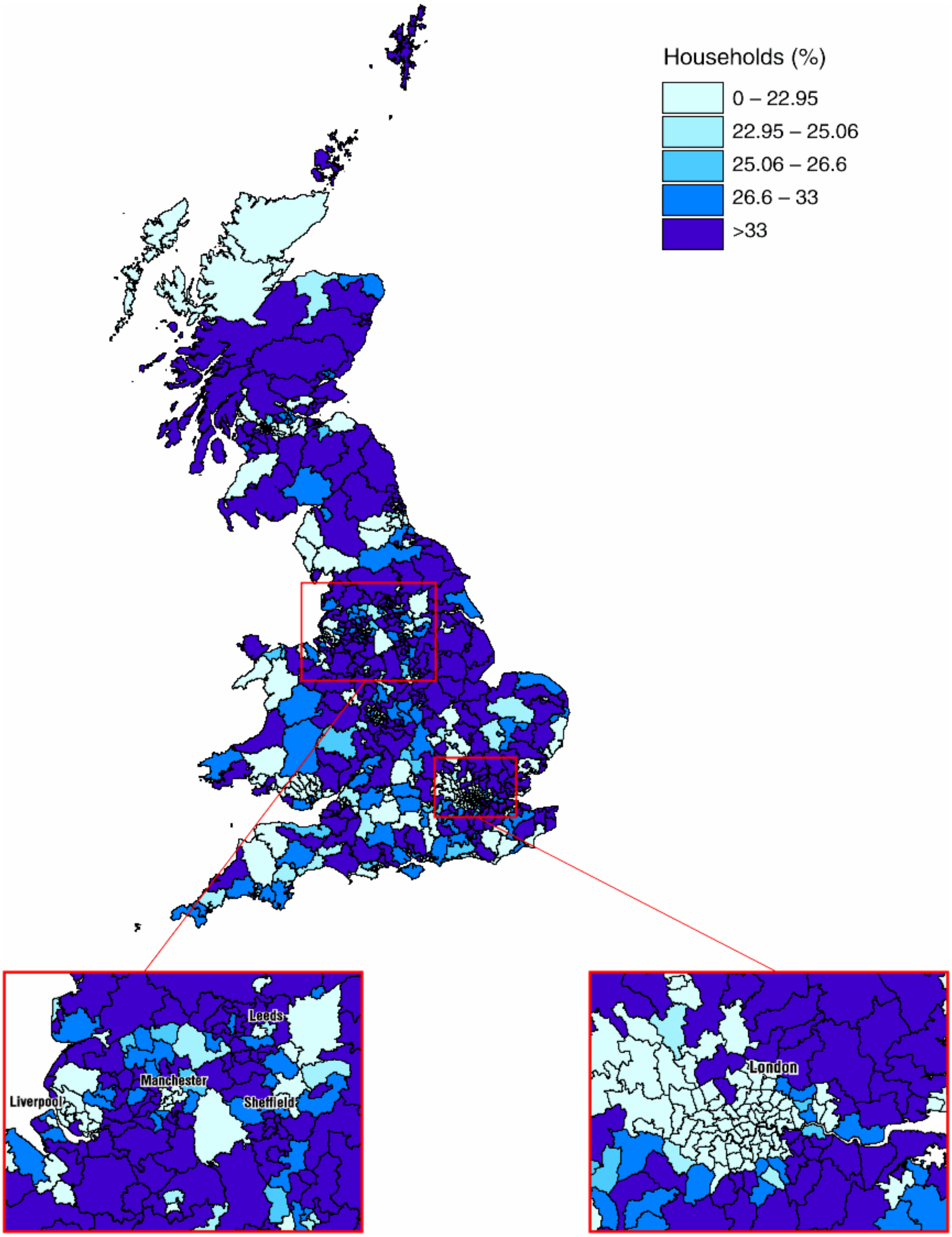


Figure 15 Projected spatial distribution of households with a retired head in 2021



11 Projecting small area microdata into the future

Once future small area statistics are available, it is then possible to apply the method described in Chapter 9 on the basis of the estimated data. This chapter presents how this can be done with an illustrative example for the City of York.

Reweighting the BHPS to fit small area population statistics for York in 1991–2021

The small area population data projection methods that were described in the previous chapter were used here for the City of York. Figures 16–20 describe the trends that were projected on the basis of past trends, using the method which was described in the previous chapter.

As can be seen in Figure 16, there is a projected significant decline in the proportions of married couples and a respective increase in the proportion of other types of households. Also, the number of lone-parent households as a proportion of all households in York is projected to remain stable at around 10 per cent.

Figure 17 describes some trends in proportions of different household classes by occupation in York. As can be seen, there is a projected polarisation. In particular, the number of the highest status households in York is projected to rise to 17.2 per cent in 2021 from 12 per cent in 1991. Likewise, the number of the lowest status households is projected to rise to 27.5 per cent of the total households from 21.2 per cent in 1991. On the other hand, the number of middle status households is projected to fall to 19.5 per cent in 2021 from 25.2 per cent in 1991. Further, there is a projected increase in the proportions of households with a retired head and a decrease in the proportion of households with an economically inactive but not retired head of household.

Figure 18 shows the trends in household car ownership in York, throughout the simulation period. There is a significant projected increase in the proportion of households that have two or more cars (projected to reach 40 per cent of all households by 2021 from just above 10 per cent

Figure 16 Trends in household structure in York, 1991–2021

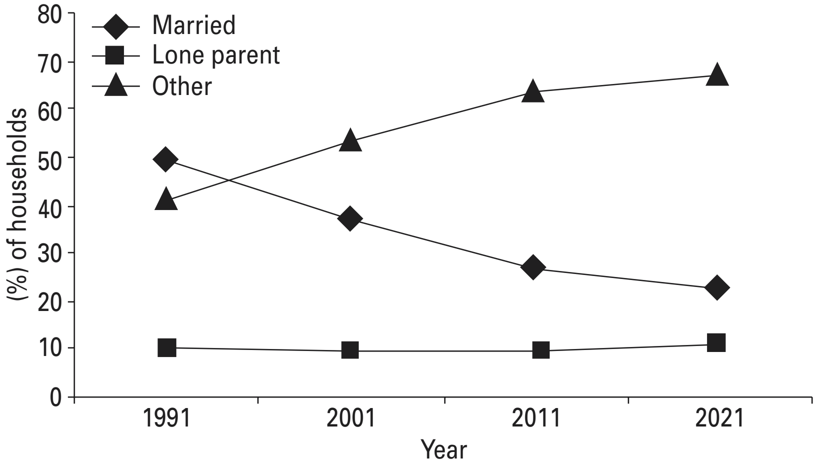


Figure 17 Trends in household class structure in York, 1991–2021

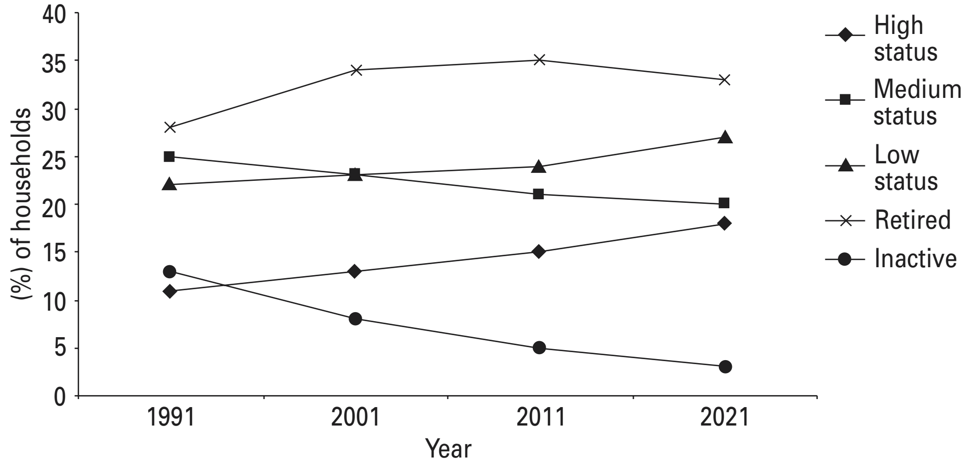
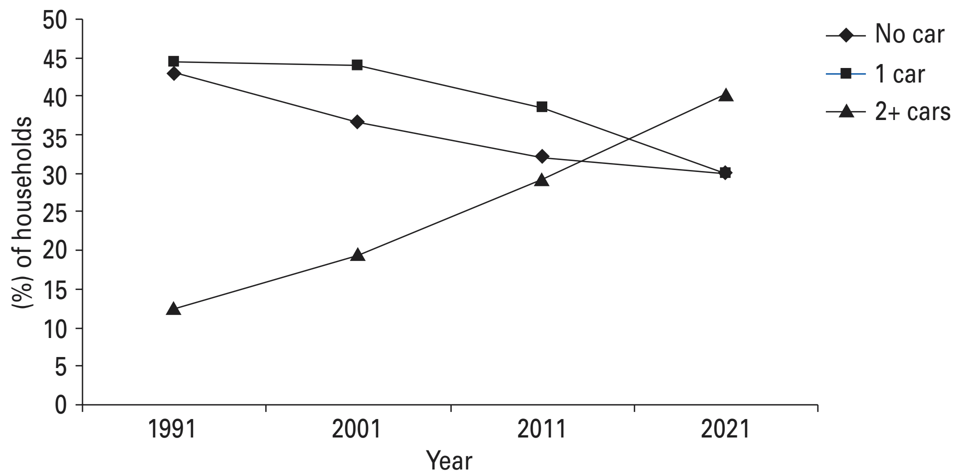


Figure 18 Trends in household car ownership in York, 1991–2021



in 1991), whereas there are decreasing trends in the numbers of households with no car or with only one car.

Figure 19 shows the changing trends in household tenure. There is a projected decrease in the proportion of owner-occupiers. Likewise, there is a significant decline in the proportion of households who live in council houses and an increase in the proportion of households living in rented accommodation.

Figure 20 shows some demographic trends in York. There is an increasing trend in the proportions of households without children, which are projected to comprise 83.2 per cent of all households in York (from 75.9 per cent in 1991). In contrast, the proportions of households with one child and with two or more children are projected to decrease slightly throughout the simulation period.

Having projected the socio-economic data into 2001, 2011 and 2021, it is possible to apply the methodology described in Chapter 9 to

Figure 19 Trends in household tenure in York, 1991–2021

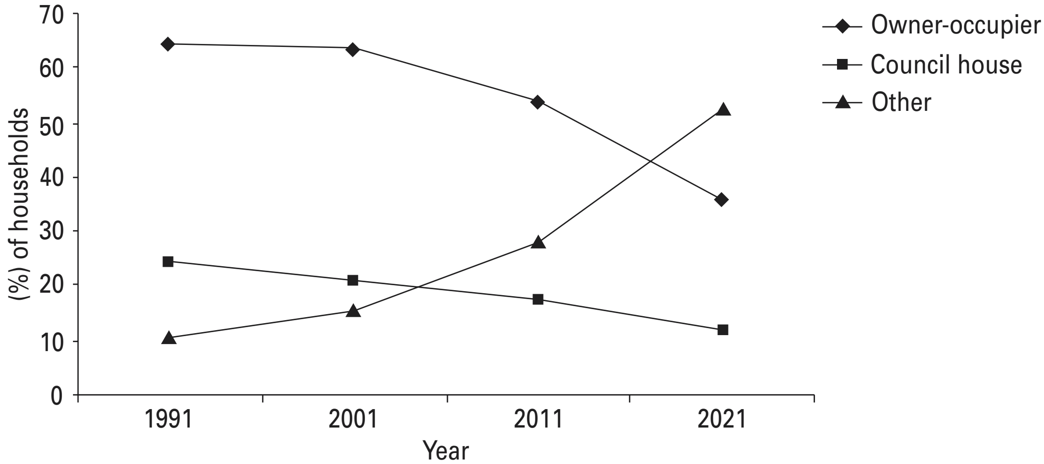
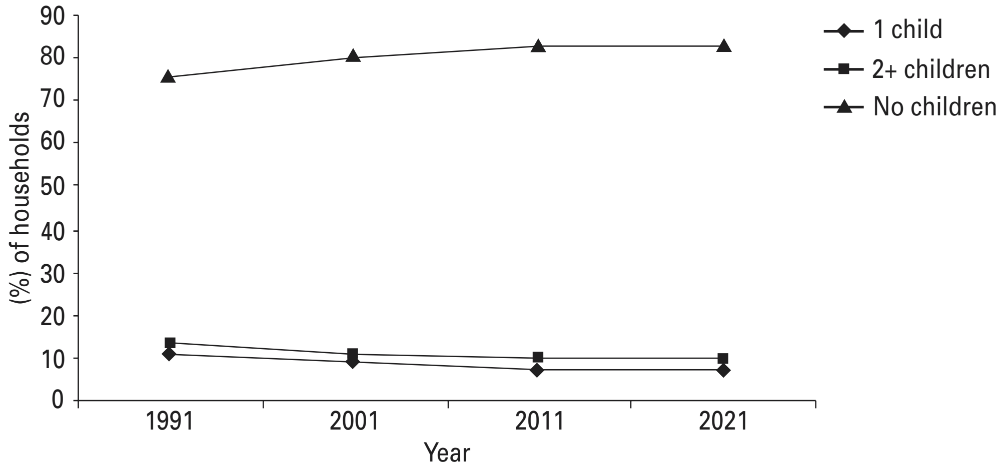


Figure 20 Number of children in households in York, 1991–2021



readjust the weights of surveyed households to fit future data. It is then possible to examine future distributions of variables on the basis of the combination of the small area data projections and the microdata. For example, the weights of 1991 BHPS households can be readjusted to represent the projected picture of York in 2001, 2011 and 2021. Such an exercise adds value to projections such as those described in Figures 16–20. It complements them with additional estimates of all the variables that are included in the BHPS. At this stage, it has to be stressed that the reliability of these estimates depends on the correlation of the survey variables with the small area projected variables. For example, the BHPS has a variable, which may be described as an index of 'loneliness', which may be related to a number of variables measured from the census at the small area level. In particular, the BHPS asks the following question to all surveyed individuals:

Is there anyone who you can really count on to listen to you when you need to talk?

The possible answers to this question are as follows:

- *yes one person*
- *yes more than one person*
- *no one.*

The way in which this question is answered will depend to a certain degree on the individual's circumstances and characteristics. For instance, it may depend on the type of household to which the individual belongs. A reasonable argument may be that people living on their own are more likely to answer 'no one' to the above question, when compared with people living in households comprising two persons or more. This relationship may well change over time and that would not be projected by the model presented here (but see Chapter 12). It is also interesting to note that, as seen in Figure 16, there is a projected increase in the numbers of households who are not 'married' or are 'lone parent'. These include single-person households. Figure 21 shows a selection of estimated trends based on the reweighting of the BHPS 1991 wave to fit the York projections (for the variables described in Figures 16–20).

As can be seen, the percentage of individuals who, according to the simulation, would feel that they would have no one to listen to them if they needed to talk is estimated to increase between 1991 and 2021.

One of the particular strengths of the method presented here is that it enables the small area estimation of survey variables on the basis of a combination of survey data and small area statistics (whether actual or projected). One of the most important and policy-relevant variables is household income. It can be argued that income is one of the main determinants of the standard of living of an individual or a household. Figure 22 shows the trends in average household income in York.

There is a slight increase in the level of average household income in the period

Figure 21 Simulated variables in York, 1991–2021

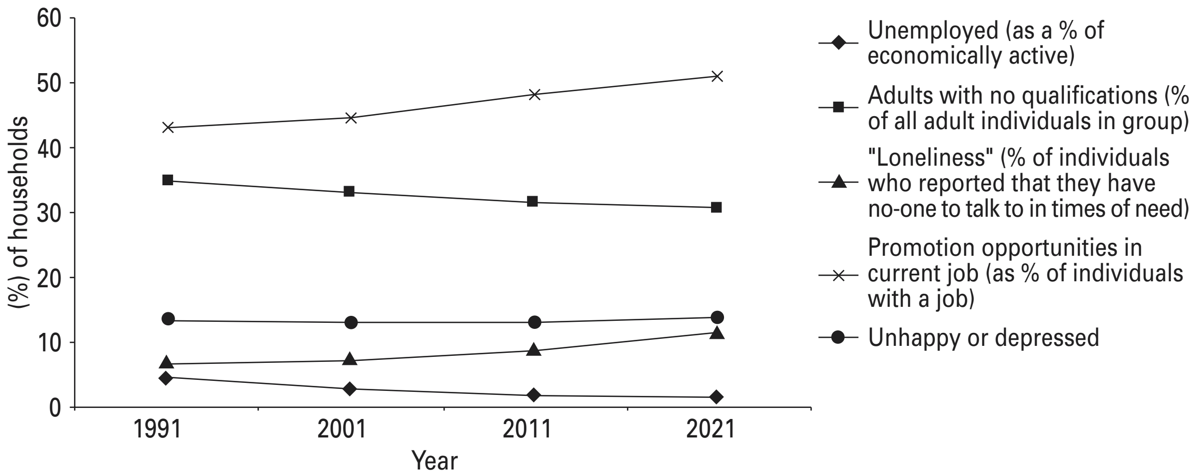
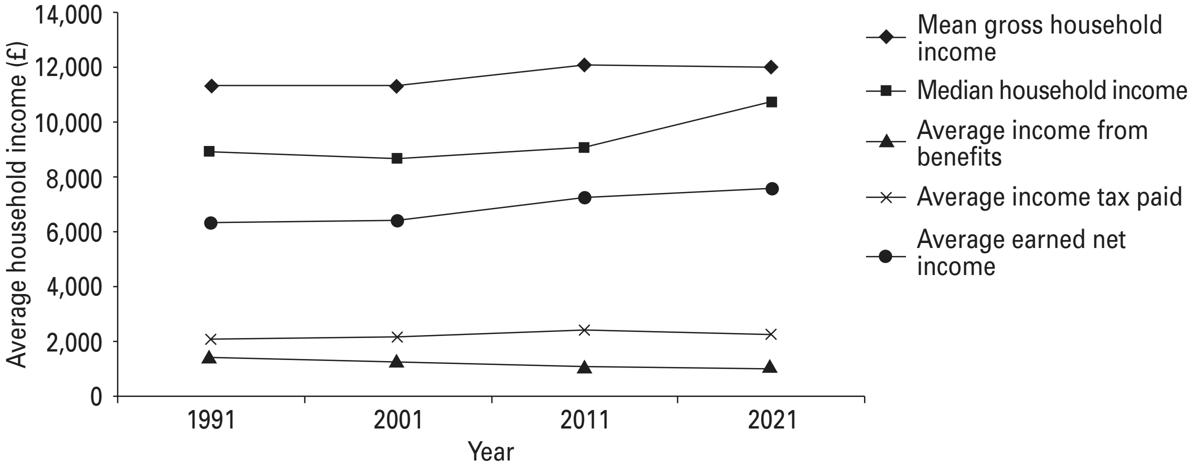


Figure 22 Simulated income trends in York, 1991–2021 (figures based on 1991 prices)



1991–2011, whereas there is a small decrease in the period 2011–21. The income inflation and deflation reflects simulated socio-economic change and it has not been altered in any way to reflect actual or predicted inflationary trends (the figures used here refer to 1991 incomes in pounds). So this income change reflects the predicted real economic growth or shrinkage in the simulated period.

Strengths and limitations

This chapter gave an illustration of how the geographical microsimulation method described in Chapters 9 and 10 can be applied to real data to provide a wide range of estimated and projected data. The main strength of this approach is the ability to add geography to survey data, or, to put it in another way, to add value to small area data (whether actual census data or projected data). Nevertheless, the main drawback of the method is that the reliability of the projected survey-based information will depend on the correlation of the survey variable with the small area measured variables (in the example given here, the variables were: *household type*, *number of children*, *number of cars*, *socio-economic group* and *tenure*). In addition, variables such as household income are affected by inflation and cannot be adequately estimated with the method presented here. Nevertheless, it can be argued that it may be useful to perform income analysis without allowing for inflation, as this may give a reflection of real and not nominal growth. The next chapter discusses ways of making the method more dynamic.

12 Projecting detailed survey data into the future

The method described in the previous chapters is capable of generating a simulated population at several points in time. It is dynamic in the sense that it can generate estimated future population microdata on the basis of past trends. Nevertheless, it does not exploit the panel nature of the BHPS, as it uses data from only one wave. The method does not attempt to use the life history data to predict what will happen in the life of the simulated individuals and households in the future. One of the reasons for not attempting to carry out such a task was that, as pointed out in Chapter 2, behavioural modelling is an extremely difficult process. Further, there has been a lack of sufficiently good migration data that would enable migration of individual and households to be accurately predicted. This chapter presents some alternative ways of making this method more dynamic by taking advantage of the BHPS life history data. In particular, it presents a novel method of simulating life histories and shows how this method has been implemented for Wales (this method required a much larger area than the district level and hence the focus moved from York to Wales).

Background

In order to exploit the panel nature of the BHPS, a method of creating household histories that could be used to populate areas was developed. As is the case with the methods presented so far in this book, the database has to be relatively simple in concept: a row for each household and separate columns for each year for which data exist. A third dimension would contain whatever variables of interest were selected. When a household ceased to exist, the row would cease and, when a new household formed, a new row would be created. Insofar as the composition of the country was changing, this would be reflected in the varying proportions of households of different types present in the database.

The task of creating these histories raises a critical question of definition. The standard definition of a household used in the BHPS and the census is cross-sectional: the sharing of facilities at a given point in time. However, there are arguments about what constitutes a 'longitudinal' household, given that the address or the composition of a household might change over time. As such, the question arises whether a household remains the same unit, given that individual household members may have left, because of death or moving out. By contrast, new household members may have entered the household, such as a new birth or persons moving in.

One of the most extensive discussions of operational longitudinal household definition is provided by Ernst *et al.* (1984). In particular, they concentrate on definitions in terms of what they call 'Same Householder', where a household is defined by reference to the head of household, and 'Reciprocal Majority', where a household is defined by reference to the proportion of household members who are present in the household at time t and at time $t + 1$. The difficulties with the first approach relate to the definition of the head of household, which for most statistical purposes pays greater attention to economic activity than household continuity; and with the second that it can produce somewhat arbitrary results depending on the size of the household and the time period employed.

An alternative approach is provided by Frick and Haisken-DeNew (2001) of the German Socio-economic Panel. They concentrate on change of address as the critical factor in identifying household dissolution and formation. Specifically, new households evolve when one or more individuals leave a pre-existing household and become resident at a new address. The drawback of this approach is that, as the authors recognise:

> *… after several years a 'household' might consist of totally different persons than in the first wave.*

However, the identification of change of address as one element in a possible definition leads to a focus on the spatial dimension. Much of the literature reviewed had as its fundamental concern the dynamics of household formation and change. The perspective here is somewhat different. The concern is not so much with household processes *per se*, rather with their effects on the composition of local areas through time. If the emphasis is shifted from the household to the space in which that household lives, then a simpler approach can be adopted.

Specifically, the result is a database not of *household* histories, but rather of *household space* histories. As soon as the emphasis is shifted from the *social* to the *physical*, the problem becomes much easier to conceptualise. Thirty-year histories for a set of generic household spaces – while these are not identifiable dwellings, they can effectively be thought of as such, and over time they may see many changes in household composition. In this chapter, an illustrative example of using this method in Wales is given.

The creation of GHOSTs

The first stage is to define a GHOST – it is a *g*eneric *ho*usehold *s*pace through *t*ime. It is generic, in that individual dwellings are not differentiated, but rather dwelling types; it concentrates on household spaces, defined as the physical spaces within which household units live; and it is considered through time (30 years in this case).

How are these GHOSTs created? The starting point is the BHPS 1991–2000. This contains information on all the households interviewed over that period of time. Unfortunately, it does not identify the buildings in which these people live (while there are person identifiers and household identifiers, there is no dwelling identifier). However, it does contain information on various characteristics of the property (notably tenure, number of rooms and dwelling type) as well as the local authority area and whether individual members of the household have moved in the last year.

Individual interview histories

As a test of this approach, an attempt was made to use this information from the BHPS to create a set of GHOSTs for Wales to cover the period 1991–2020. The first step is to identify all senior household members who at any time during the period were interviewed in Wales. The household response for each year (BHPS variable *xhvfio*) is established and a record is created for each household reference person (BHPS variable *xhgr2r* = 1) or spouse (BHPS variable *xhgr2r* = 2) or partner (BHPS variable *xhgr2r* = 3) or other person recorded as head of household (BHPS variable *xhoh* = 1). For each of the above individuals interviewed, a full history of their interviews, whether in Wales or not, and of the households to which they belonged at the time is extracted.

Table 9 lists a sample of 20 such individuals together with their *personal identifying number* (*pid*) and their *household identifying number* (*hid*) at each wave. It also contains a newly derived variable called OSH or *o*riginal *s*ample *h*ousehold. This is similar to the BHPS concept of OSM (*o*riginal *s*ample *m*ember) in that every person interviewed in the BHPS is interviewed because they themselves were interviewed in the first wave (they are OSMs) or because they are currently in the same household as an OSM. The OSH is simply the first BHPS wave household identifier (BHPS variable *ahid*) for that OSM. When sorting out household histories, the OSH is an invaluable indicator because it helps to disentangle some of the more complicated household histories (for example, where members of families split off from original households only to rejoin at a later date) as well as clarifying apparent discontinuities in histories (for example, where the family reference person changes between times despite no change in household composition, or where it is the result of the gain or loss of a member of another generation).

Some of the cells in Table 9 are shaded grey: this indicates either (a) that although the individual was interviewed in the identified household, (s)he was not the head of household, household reference person (*hrp*) or spouse/partner at that time – e.g. *pid* 17942497 in wave c; or (b) although an attempt was made to interview the person, this was not successful – e.g. *pid* 17931304 in wave d. In either case, the household identifier appears as a negative integer. Where no number appears in a cell this is because the individual did not appear on the interview schedules: either (a) because the individual had been dropped by the BHPS (because of death, emigration, repeated refusal, etc.) – e.g. *pid* 17935741 from wave d onwards; or (b) because (s)he was not an OSM and was not at that time living with an OSM – e.g. *pid* 28126343 in wave a.

Household interview histories

Having extracted these individual interview histories, the next step is to convert them to household interview histories. As discussed above, a rigorous definition of household is not necessary: the ultimate aim is to create household space rather than household histories. However this methodology is effectively based on the 'Same Householder' approach described by Ernst *et al.* (1984), though with a (necessarily) flexible approach to the identification of the appropriate householder. Table 10 shows how this is undertaken, with nine (colour-coded) types of household history distinguished, together with accompanying data on interview response, relationship and change of address. The first column is the OSH: households with a common OSH are grouped within bold horizontal lines. More than one household may be identifiable within these groups and each household is itself delineated in bold.

Table 9 Individual interview histories for 20 Welsh members of the BHPS

OSH	*ahid*	*bhid*	*chid*	*dhid*	*ehid*	*fhid*	*ghid*	*hhid*	*ihid*	*jhid*	*pid*
1660683	1660683	2853957	3773728	4598016	5538963	6497748	7376642	8324875	9134948	10125833	17926793
1660772	1660772	2854171	3773906	4598202	5539153	6497926	7376804	8325022	9135014	10125906	17927919
1660861	1660861	2854317	3774023	4598326	5539277	6498043	7376901	8325138	9135065	10125949	17929067
1660969	1660969	2854457	3774147	4598458	-5539331	6498108	7376952	8325189	-9135081	-10125965	17930189
1661051	1661051	2854759	3774392	-4598695	-5539587	-6498353	–	–	–	–	17931304
1661159	1661159	2855038	3774643	4598881	5539706	6498477	7377169	8325332	9135154	10126023	17932386
1661248	1661248	2855321	3774899	4599128	5539951	6498728	7377363	8325545	9284575	10258574	17933536
1661248	-1661248	-2855321	-3774899	-4599128	-5539951	-6498787	-7377525	8325596	9299203	10268561	17933501
1661426	1661426	-2855534	-3775089	–	–	–	–	–	–	–	17935741
1661426	-1661426	2855674	3775321	4719638	5864887	6775306	7575599	8492948	9202455	10188274	17935776
1661426	–	2855607	3775143	-4719506	–	–	–	–	–	–	28126343
1661523	1661523	2855747	3775704	4599365	5540208	6498965	7377576	8325758	9135235	10126104	17936861
1661612	1661612	2855895	3775828	4599497	5540321	6499082	7377673	8325855	9135278	10126155	17937981
1661892	1661892	2856034	3775941	4599616	5540453	6499201	7377789	8325952	9135316	10126198	17941342
1661981	1661981	-2856107	3776018	-4599675	-5540518	-6499279	-7377835	-8326002	9135332	-10126228	17942462
1661981	-1661981	2856107	-3776018	4599675	5540518	6499279	7377835	8326002	-9135332	10126228	17942497
1662082	1662082	2856395	3776263	4599926	5540763	6499511	7378033	8326215	9135413	10126317	17943582
1662171	1662171	2856468	3776328	4599985	5540828	6499589	7378084	8326266	9135448	10126333	17944732
1662279	1662279	2856611	3776441	4600118	5540941	6499708	7378149	8326312	9135464	10126384	17945828
1662368	1662368	2856751	3776638	4600231	-5541077	6499821	7378254	-8326428	-9135529	-10126422	17946948

Table 10 Household interview histories derived from Table 9

OSH	*ahid*	*bhid*	*chid*	*dhid*	*ehid*	*fhid*	*ghid*	*hhid*	*ihid*	*jhid*
1660683	1660683	2853957	3773728	4598016	5538963	6497748	7376642	8324875	9134948	10125833
1660772	1660772	2854171	3773906	4598202	5539153	6497926	7376804	8325022	9135014	10125906
1660861	1660861	2854317	3774023	4598326	5539277	6498043	7376901	8325138	9135065	10125949
1660969	1660969	2854457	3774147	4598458	-5539331	6498108	7376952	8325189	-9135081	-10125965
1661051	1661051	2854759	3774392	-4598695	-5539587	-6498353	–	–	–	–
1661159	1661159	2855038	3774643	4598881	5539706	6498477	7377169	8325332	9135154	10126023
1661248	1661248	2855321	3774899	4599128	5539951	6498728	7377363	8325545	9284575	10258574
1661248	-1661248	-2855321	-3774899	-4599128	-5539951	-6498787	-7377525	8325596	9299203	10268561
1661426	1661426	-2855534	-3775089	–	–	–	–	–	–	–
1661426	-1661426	2855674	3775321	4719638	5864887	6775306	7575599	8492948	9202455	10188274
1661426	–	2855607	3775143	-4719506	–	–	–	–	–	–
1661523	1661523	2855747	3775704	4599365	5540208	6498965	7377576	8325758	9135235	10126104
1661612	1661612	2855895	3775828	4599497	5540321	6499082	7377673	8325855	9135278	10126155
1661892	1661892	2856034	3775941	4599616	5540453	6499201	7377789	8325952	9135316	10126198
1661981	1661981	-2856107	3776018	-4599675	-5540518	-6499279	-7377835	-8326002	9135332	-10126228
1661981	-1661981	2856107	-3776018	4599675	5540518	6499279	7377835	8326002	-9135332	10126228
1662082	1662082	2856395	3776263	4599926	5540763	6499511	7378033	8326215	9135413	10126317
1662171	1662171	2856468	3776328	4599985	5540828	6499589	7378084	8326266	9135448	10126333
1662279	1662279	2856611	3776441	4600118	5540941	6499708	7378149	8326312	9135464	10126384
1662368	1662368	2856751	3776638	4600231	-5541077	6499821	7378254	-8326428	-9135529	-10126422

(Continued)

Table 10 Household interview histories derived from Table 9 (Continued)

pid	*aiv*	*biv*	*civ*	*div*	*eiv*	*fiv*	*giv*	*hiv*	*iiv*	*jiv*
17926793	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01
17927919	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.02	10.01
17929067	10.02	10.02	10.02	10.01	10.01	10.01	10.01	10.01	10.01	10.01
17930189	10.02	11.02	-12.02	12.02	16.02	10.02	10.01	10.01	61.00	16.01
17931304	10.02	11.02	12.01	60.00	61.00	92.00	–	–	–	–
17932386	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01
17933536	10.03	10.03	10.03	10.02	10.03	10.01	10.01	-10.01	10.01	10.01
17933501	10.01	10.01	10.01	10.01	10.01	50.00	50.00	-10.01	10.01	10.01
17935741	12.01	61.00	92.00	–	–	–	–	–	–	–
17935776	12.23	-10.02	10.01	10.01	-10.01	10.01	10.01	10.01	10.01	10.01
28126343	–	10.01	10.01	91.00	–	–	–	–	–	–
17936861	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01
17937981	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01
17941342	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01
17942462	10.01	10.15	10.01	11.13	11.15	12.13	12.15	12.15	12.01	12.15
17942497	10.04	10.02	10.04	11.01	11.02	12.01	12.02	12.02	12.04	12.02
17943582	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01	10.01
17944732	10.02	10.02	10.02	10.02	10.02	10.01	10.01	10.01	12.01	12.02
17945828	10.01	10.01	10.01	10.01	10.01	10.01	10.02	10.01	10.02	10.01
17946948	10.01	10.01	10.01	10.01	16.01	10.01	10.01	50.00	50.00	50.00

(Continued overleaf)

Table 10 Household interview histories derived from Table 9 (Continued)

ahid–jhid	Interpretation
1660683	Household identifiable for ten years with maximum one year's missing data
2865947	Household identifiable for ten years but treated in parts because of missing data
2855674	Newly formed household with history through to 2000
6640567	Short-term household: OSM respondent temporarily in another household
3781577	Short-term household: non-OSM respondent housing 'dependent' OSM
4833333	Household identifiable until respondent moves into institution
1664697	Household identifiable until respondent's death
1673653	Household identifiable until respondent moves out of scope
1660969	Household history terminated because of non-response

Integer part of *aiv–jiv* (subset)	Decimal part of *aiv–jiv* (subset)
10 Full household interview	0.00 Not applicable
11 >= 1 interview; >= 1 proxy	0.01 Household reference person
12 >= 1 interview; >= 1 refusal	0.02 Spouse
16 Telephone proxy only	0.03 Partner
50 Address not found	0.04 Child
60 Refusal to research centre	0.13 Parent
61 Refusal to interviewer	0.15 Parent-in-law
91 No OSM present	0.23 Lodger
92 Adamant refusal	

Each successful interview of a Welsh household is represented by a coloured cell containing the household identifier in columns *ahid–jhid*. The key to the various possible colourings is shown below Table 10. A negative value in any of these columns indicates either that another respondent in the same household is responsible for a coloured cell (and hence 'represents' that household) or that a successful interview was not conducted or that it took place outside Wales.

Columns *aiv–jiv* contain the household interview result (BHPS variable *xhvfio*) as the integer and the relationship of the respondent to the household reference person (*xhgr2r*) as the decimal part. Negative values indicate that the respondent moved in the last year, in which case a broken line also appears between the years in question in columns *ahid–jhid*. An integer value of 99 signifies a respondent living outside Wales.

The real value of these data will become apparent when household space histories are created, but Table 10 also demonstrates some of the difficulties in assembling 'simple' household histories from panel data such as the BHPS. For example, OSH 1661981 did not move during the ten years of the study, but did contain varying numbers of people throughout that time. Furthermore, two individuals were identified as household reference person at different times and with no apparent consistency. Thus parent (*pid* 17942462) was accorded that status in waves a, c and i, whereas child (*pid* 17942497) was *HRP* in the remaining waves. By organising the data as shown in Table 10, it becomes possible to identify the continuity that exists beneath apparent instability and underlines some of the potential pitfalls facing a mechanistic implementation of the 'Same Householder' approach.

Household space histories

The next step is to convert these household interview histories into household space histories, a process that is achieved in two stages. The first part of the process, represented in Table 11, involves the separation and reordering of the various individual household histories identified above. First, any household that was identified in all waves of the study, and that did not move address during those ten years, is accorded a single row. Second, any household with a full history but that did move is listed in consecutive rows with a new row for each new address. And, third, are the remaining interview responses, with a row for each.

In terms of the 20 histories already identified in Tables 9 and 10, this produces: first, the 11 continuous household histories that did not involve a change of address (which in the case of OSH 1661981 involves the collapsing of two rows from Table 9 to just one in Table 11); second, the sole example of a continuous history incorporating one change of address, producing two rows in Table 11; and, third, the seven remaining rows from Table 10, which produce nine rows in Table 11 because of the two changes of address affecting *chid* 3774147 and *ehid* 5864887 respectively. In terms of the entire Welsh subset of the BHPS, this approach distinguishes 140 households that were identified for ten years and that did not move during that time; 45 households (covering 102 rows) that have a full history but that did move; and the remaining interview responses that between them cover over 800 cells, equivalent to more than 80 ten-year household histories.

Table 11 Individual household space histories derived from Table 9

ahid	*bhid*	*chid*	*dhid*	*ehid*	*fhid*	*ghid*	*hhid*	*ihid*	*jhid*
1660683	2853957	3773728	4598016	5538963	6497748	7376642	8324875	9134948	10125833
1660772	2854171	3773906	4598202	5539153	6497926	7376804	8325022	9135014	10125906
1660861	2854317	3774023	4598326	5539277	6498043	7376901	8325138	9135065	10125949
1661159	2855038	3774643	4598881	5539706	6498477	7377169	8325332	9135154	10126023
1661523	2855747	3775704	4599365	5540208	6498965	7377576	8325758	9135235	10126104
1661612	2855895	3775828	4599497	5540321	6499082	7377673	8325855	9135278	10126155
1661892	2856034	3775941	4599616	5540453	6499201	7377789	8325952	9135316	10126198
1661981	2856107	3776018	4599675	5540518	6499279	7377835	8326002	9135332	10126228
1662082	2856395	3776263	4599926	5540763	6499511	7378033	8326215	9135413	10126317
1662171	2856468	3776328	4599985	5540828	6499589	7378084	8326266	9135448	10126333
1662279	2856611	3776441	4600118	5540941	6499708	7378149	8326312	9135464	10126384
1661248	2855321	3774899	4599128	5539951	6498728	7377363			
							8325545	9284575	10258574
1660969	2854457								
		3774147	4598458	-5539331	6498108	7376952	8325189		
1661051	2854759	3774392							
							8325596	9299203	10268561
1661426									
	2855674	3775321	4719638						
				5864887	6775306	7575599	8492948	9202455	10188274
	2855607	3775143							
1662368	2856751	3776638	4600231	-5541077	6499821	7378254			

The second part of the process, represented in Table 12, involves the compression and combination of information from the movers to create virtual histories for those household spaces that at the moment are represented by fragments. The first 11 rows of Table 12 are the same as the corresponding rows of Table 11, but also contain ten additional variables, which provide information on the dwelling in which the interview took place. It is worth noting that, despite the fact that none of these households reported moving, there are not infrequent changes in household space information (which, as the accompanying key explains, consists of information on tenure, dwelling size and type). Some of this will be real (changes in tenure under right to buy or extensions, for example), but the remainder must reflect coding/response errors either in the household space information or else in the question on moves. For present purposes, any apparent changes in household space information have been ignored and it has been assumed that the question on moves was correctly answered.

Variables *ahid–ghid* and *ahs–ghs* of the next row reflect the actual history of OSH 1661248. As is shown in Table 10, *pid* 17933536 moved between waves g and h (as Table 10 shows, from a semi-detached to a terraced house) and the virtual history for the semi is completed by the history of another household that moved into just such a dwelling in wave *h*. The history is virtual, in that the chances are infinitesimally small that it is the same dwelling, but there is a perfect match on tenure, dwelling type and size. The remainder of *pid* 17933536's history appears in the next row, where it complements that of another Welsh household that lived in a terraced house until moving between waves g and h.

Table 12 provides five further examples of perfect matching on household space variables, each distinguished by green shading of variables *ahs–jhs*. The remaining five rows are at least 50 per cent unshaded, but with varying degrees of strong colour shading. In all but one of these cases the colour in question is yellow, used to denote a close, but not perfect, match on household space variables at the interface between the unshaded and shaded cells. Thus, in the first such example in Table 12, an otherwise perfectly matched history is completed by the incorporation of one year's data from a detached, rather than a semi-detached, dwelling, in both cases privately rented with four/five rooms. This is a perfectly acceptable compromise. And in the one instance, distinguished by blue shading (primarily used to denote matching across different categories of tenure), again this is acceptable in order to maximise the information available.

In terms of the entire Welsh subset of the BHPS, in addition to the 140 households that were identified for ten years and that did not move during that time, there are: a further 66 rows that have been produced by perfect matching of household space descriptors; another 38 rows where the match is perfect on tenure but where there is a small difference in one of the other elements – a semi rather than a detached house or a small one rather than a medium-sized one; and 11 further rows where greater liberty has been taken with matching (but where two-thirds or more of the history for each row is consistent and whose inclusion helps to redress what would otherwise be a bias against rented tenures).

Table 12 Generic household space histories derived from Table 11

ahid	*bhid*	*chid*	*dhid*	*ehid*	*fhid*	*ghid*	*hhid*	*ihid*	*jhid*
1660683	2853957	3773728	4598016	5538963	6497748	7376642	8324875	9134948	10125833
1660772	2854171	3773906	4598202	5539153	6497926	7376804	8325022	9135014	10125906
1660861	2854317	3774023	4598326	5539277	6498043	7376901	8325138	9135065	10125949
1661159	2855038	3774643	4598881	5539706	6498477	7377169	8325332	9135154	10126023
1661523	2855747	3775704	4599365	5540208	6498965	7377576	8325758	9135235	10126104
1661612	2855895	3775828	4599497	5540321	6499082	7377673	8325855	9135278	10126155
1661892	2856034	3775941	4599616	5540453	6499201	7377789	8325952	9135316	10126198
1661981	2856107	3776018	4599675	5540518	6499279	7377835	8326002	9135332	10126228
1662082	2856395	3776263	4599926	5540763	6499511	7378033	8326215	9135413	10126317
1662171	2856468	3776328	4599985	5540828	6499589	7378084	8326266	9135448	10126333
1662279	2856611	3776441	4600118	5540941	6499708	7378149	8326312	9135464	10126384
1661248	2855321	3774899	4599128	5539951	6498728	7377363	8630593	9292322	10264159
1682512	2882175	3799131	4619153	5557518	6515541	7390025	8325545	9284575	10258574
1660969	2854457	3796787	4803841	5562899	6858333	7625022	8530238	9216928	10200703
1675427	2871386	3774147	4598458	-5539331	6498108	7376952	8325189	9185771	10302174
1661051	2854759	3774392	4616685	5872448	6781985	7580509	8636788	9282688	10292594
1664603	2859661	3779211	4602099	5542502	6501133	7379358	8325596	9299203	10268561
1661426	2870967	3789365	4804457	5804531	6732429	7549512	8471509	9194371	10180834
1671596	2855674	3775321	4719638	5881129	6789641	7601328	8623422	9310134	10307494
1687743	2890461	3806375	4625099	5864887	6775306	7575599	8492948	9202455	10188274
1664883	2855607	3775143	4813626	5844975	6837638	7676921	8570132	9232133	10213554
1662368	2856751	3776638	4600231	-5541077	6499821	7378254	8496315	9287655	10260803

(Continued)

Table 12 Generic household space histories derived from Table 11 (Continued)

ahs	*bhs*	*chs*	*dhs*	*ehs*	*fhs*	*ghs*	*hhs*	*his*	*jhs*
132	132	132	121	132	122	132	132	112	122
121	121	121	121	121	122	121	131	121	121
222	222	222	222	222	222	222	222	222	222
121	121	121	121	121	121	121	121	121	121
121	121	121	121	120	121	120	121	121	121
130	131	131	131	131	131	131	131	131	131
121	122	122	122	122	122	122	122	121	122
122	122	122	122	122	122	122	122	122	122
111	114	111	111	111	111	111	111	111	111
121	122	121	121	111	122	121	121	121	122
120	121	121	121	121	121	121	121	121	121
122	122	122	122	122	122	122	122	122	122
123	123	123	123	123	123	123	123	123	123
223	223	223	322	321	322	322	322	322	322
131	131	131	131	0	131	131	131	131	131
222	232	222	212	212	222	222	222	223	323
222	212	212	222	222	212	212	212	212	212
222	223	223	223	223	223	223	223	223	223
331	331	321	311	312	322	322	322	322	322
222	222	222	222	222	222	222	222	222	222
122	122	112	131	131	113	112	113	113	113
120	121	121	111	0	111	111	111	111	111

(Continued overleaf)

Table 12 Generic household space histories derived from Table 11 (Continued)

122	Same household resident throughout
123	Perfect match of household space variables at changeover
131	Household space variables for most of period
312	Close match of household space variables at changeover
114	Tenuous but acceptable match of household space variables at changeover

First digit	Second digit	Third digit
0 No information	0 No information	0 No information
1 Owner-occupied	1 < = 3 rooms	1 Detached
2 Social rented	2 4/5 rooms	2 Semi-detached
3 Other rented	3 > = 6 rooms	3 Terraced
4 Flat or maisonette		

This leaves some 140 cells (approximately 5 per cent of the total valid interview record for Welsh households in the BHPS) that could not be matched without overstretching the bounds of credulity. As a result of this flexible matching strategy, the sample contains a changing profile of household spaces that (more or less) matches the changing profile of Welsh household spaces over the ten years in question.

Projecting household spaces through time

There are now 255 household space histories covering the period 1991–2000. They are generic, in that 115 of them do not refer to the same physical dwelling, but each contains contemporaneous information. The next task is to project these histories through time: to populate these 255 household spaces through to the year 2020.

To do this, the known histories for 1991–2000 are replicated, twice, each time for ten years. Thus the household characteristics for household space *x* in 2005 will be based on that from a known household in 1995; that for 2010 will be based on the same household space's history in 2000, etc. Without some form of updating methodology, this is self-evidently unrealistic, but this is only part of a two-part process. Once the GHOSTs are in place, then a methodology must be provided for dealing with change in variables beyond 2000. Insofar as these changes are a function of population composition, then the fact that these simulations are dependent on external 'compositional' constraints should be sufficient.

As with the conversion of household to household space histories, the projections are done in stages. In the first stage, represented in Table 13, the 255 GHOSTs for 1991–2000 are projected forward to 2010. First, the existing histories (from Table 12) are duplicated and copied to their right (as columns *hs01–hs10*), thereby producing twice as many columns. Then the pre-existing half of the table (columns *ahid–jhs*) is sorted by *jhs* and the new half of the table is sorted by *hs01*. Within each category of *jhs* and *hs01*, rows are further sorted by income so that the less affluent households appear first and the most affluent at the end of each respective category.

If there had been no change in the relative proportions of the different combinations of household space characteristics, then there would be an identical match between *jhs* and *hs01* for each row in the table. In the case of the sample that has been followed up to this point, shown in Table 13, this is generally true: all but three have a straight match between *jhs* and *hs01* (allowing for the presence of missing information in a few cases). All but 30 of the 255 rows are matched (denoted by coloured shading of the *jhs* and *hs01* cells). Furthermore, because of the sorting by income, similar households (as well as household spaces) are matched.

In addition, where there is an acceptable match between *jhs* and the cell from the following year (not shown in Table 13, but using analagous principles), the latter is shaded in grey.

This still leaves six cases where no acceptable match was possible. This is inevitable given the changing profile of household spaces over time; indeed, it is somewhat remarkable that as few as six out of 255 Welsh histories failed to pass the matching test. In these six cases, the 2001–10 history was produced by duplicating another record that does provide a perfect match.

The second stage, taking the projections through to 2020, is achieved in analogous fashion. In total, 226 of the 255 rows are perfectly matched, 23 are matched on a best-fit basis, while another six histories were produced by duplication. This results in the household ids for 255 Welsh GHOSTs, which can be used to populate Wales for the period 1991–2020.

A worked example

As an example of what these histories mean in practice, consider the first row of Table 13. The household that lived in this smaller than average owner-occupied detached house in 1991 (*ahid* = 1662082) comprised a single pensioner who continued to live alone in the same house right through to 2000. The successor household for *hs01* (*hid01* = 1677292) also comprised a single pensioner. However, within two years, that pensioner had married and the couple continued to live in the same house for the next six years. No interview was achieved with the couple in wave h and hence new occupants were used to fill the generic household space. The new household comprised a married couple with their three children who provide the history for *hs09* and *hs10*. The next ten years are not listed in Table 13 but the successor household for *hs11* was similar to that in *hs10*, though with one fewer child. This situation persisted up to and including *hs14*, after which one of the parents left, leaving a single-parent household with two dependent children to live in the house up to and including *hs20*.

This example serves to illustrate the important characteristics of the approach. In this case, a plausible history is generated for the household space in question. This will not always be the case, of course, with discontinuities concentrated at the changeover points between 2000/01 and 2010/11.

Table 13 GHOSTs derived from Table 12 projected through to 2010

ahid		*jhid*	*ahs*	*bhs*	*chs*	*dhs*	*ehs*	*fhs*	*ghs*	*hhs*	*his*	*jhs*
1662082	……………..	10126317	111	114	111	111	111	111	111	111	111	111
1662368	……………..	10260803	120	121	121	111	0	111	111	111	111	111
1664883	……………..	10213554	122	122	112	131	131	113	112	113	113	113
1661159	……………..	10126023	121	121	121	121	121	121	121	121	121	121
1661523	……………..	10126104	121	121	121	121	120	121	120	121	121	121
1662279	……………..	10126384	120	121	121	121	121	121	121	121	121	121
1660772	……………..	10125906	121	121	121	121	121	122	121	131	121	121
1661892	……………..	10126198	121	122	122	122	122	122	122	122	121	122
1662171	……………..	10126333	121	122	121	121	111	122	121	121	121	122
1661981	……………..	10126228	122	122	122	122	122	122	122	122	122	122
1660683	……………..	10125833	132	132	132	121	132	122	132	132	112	122
1661248	……………..	10264159	122	122	122	122	122	122	122	122	122	122
1682512	……………..	10258574	123	123	123	123	123	123	123	123	123	123
1661612	……………..	10126155	130	131	131	131	131	131	131	131	131	131
1675427	……………..	10302174	131	131	131	131	0	131	131	131	131	131
1664603	……………..	10268561	222	212	212	222	222	212	212	212	212	212
1660861	……………..	10125949	222	222	222	222	222	222	222	222	222	222
1687743	……………..	10188274	222	222	222	222	222	222	222	222	222	222
1661426	……………..	10180834	222	223	223	223	223	223	223	223	223	223
1661051	……………..	10292594	222	232	222	212	212	222	222	222	223	323
1660969	……………..	10200703	223	223	223	322	321	322	322	322	322	322
1671596	……………..	10307494	331	331	321	311	312	322	322	322	322	322

(Continued)

Table 13 GHOSTs derived from Table 12 projected through to 2010 (Continued)

hs01	*hs02*	*hs03*	*hs04*	*hs05*	*hs06*	*hs07*	*hs08*	*hs09*	*hs10*	*hid01*		*hid10*
111	111	111	111	111	121	111	111	111	111	1677292	....	10303359
111	111	121	121	121	121	121	121	121	121	1681117	....	10130055
113	113	113	123	123	133	132	122	132	122	1689797	....	10132007
121	121	122	122	121	121	122	121	121	122	1673467	....	10128271
120	121	121	121	121	121	121	121	121	121	1662279	....	10126384
121	121	121	121	121	121	121	121	121	121	1668528	....	10127445
121	121	121	121	120	121	120	121	121	121	1661523	....	10126104
122	122	122	122	122	122	122	122	122	122	1682059	....	10130306
122	122	4	121	122	121	121	122	121	121	1691848	....	10132589
122	122	122	132	122	122	132	122	122	132	1687468	....	10131388
221	122	122	122	132	122	120	122	122	122	1662732	....	10126449
122	122	122	121	121	121	121	121	121	121	1674773	....	10176845
123	123	123	123	123	123	123	123	123	123	1690728	....	10218971
131	131	131	131	131	122	131	131	131	131	1667688	....	10225935
131	121	121	121	121	121	121	121	121	0	1686801	....	-10131159
212	112	112	112	112	112	112	110	112	0	1683071	....	-10130438
222	223	223	222	223	222	222	222	222	0	1690256	....	-10132198
222	222	202	222	232	232	202	222	122	122	1664328	....	10126694
223	223	223	223	223	222	223	223	223	223	1680749	....	10129669
223	223	223	223	223	222	223	223	223	223	1680188	....	10129529
322	322	322	323	322	322	322	322	322	322	1676733	....	10128689
331	331	321	311	312	322	322	322	322	322	1671596	....	10307494

(Continued overleaf)

Nevertheless, many of the histories are credible and can be used to illustrate the nature of change affecting household spaces over time. Furthermore, the primary use of these GHOSTs is when aggregated to provide area estimates over time, at which point any inconsistencies in individual histories will be subsumed within those aggregate statistics.

A regression framework for continuous variables

Any methodology for projecting estimates needs to be able to model those estimates in terms of household characteristics (specifically in terms of the six constraint variables presented in the previous chapter) and to deal with both discrete and continuous variables. It soon became apparent that this might be achieved within a general regression framework, employing logistic regression for the binary variables. In order to test this approach, it was decided to use existing data from the BHPS to model changes in household income (strictly speaking log of household income thereby compensating for income's positively skewed distribution and minimising the biasing effect of extreme values) and personal computer (PC) ownership in Wales. If an appropriate model could be devised, then it could be used to update the GHOSTs presented above through to 2021. The Appendix discusses how it is possible to build such a model for both continuous variables (such as household income) and binary variables (such as PC ownership).

Table 13 GHOSTs derived from Table 12 projected through to 2010 (Continued)

First digit	Second digit	Third digit
0 No information	0 No information	0 No information
1 Owner-occupied	1 < = 3 rooms	1 Detached
2 Social rented	2 4/5 rooms	2 Semi-detached
3 Other rented	3 > = 6 rooms	3 Terraced
		4 Flat or maisonette

13 Assessing simulation outputs and policy relevance

As pointed out in Chapter 11, the reliability of the projected survey-based information will depend on the correlation of the survey variables with the small area measured variables (in the example given here, these variables were: *household type*, *number of children*, *number of cars*, *socio-economic group* and *tenure*). This chapter discusses ways of validating the simulation outputs, and of determining where it works best and where it performs poorly. In order to do so, actual data for Wales and York are compared to their simulation outputs. Simulated and actual data for the whole of Britain at the parliamentary constituency level are also compared. Some comparisons of actual data and official government projections with the outputs of the projection method detailed in Chapter 10 are also presented.

Validating the simulated outputs with the use of census data

The census of population provides an obvious benchmark against which to test the results from the simulation method. As illustrated in Chapter 9, a reweighting methodology was used to select the BHPS household that best matched five small area census household variables. This section presents a comparison of the simulated values with their census counterparts for the five selected variables, which were not used as constraints in the simulation. These variables are listed in Table 14 together with their average values in the BHPS and in the census.

Several points are worth noting. First, as most census variables are counts, the majority of comparisons relate to proportions. The one exception, average age, assumes an even distribution of ages within each of the census age categories but this is not unreasonable. Second, the average values of four variables match very closely and the exceptions do not differ to such a degree as to make comparisons useless. Nevertheless, it will be worth investigating why the average values for unemployment differs by 2 per cent. Third, the variables selected combine various aspects of household and individual characteristics. Age is an individual-level characteristic available for all residents including children. Illness, unemployment, education and travel to work are all measured at the individual level but relate to different sections of the adult population.

In an effort to predict how well each variable might be modelled by the method, a series of individual-level regressions was undertaken

Table 14 Comparing five variables between the 1991 census and the BHPS (wave 1) at the national level

Variable	BHPS	UK census
Average age of residents	37.6	37.5
Proportion of adults with long-term illness	0.146	0.149
Proportion of economically active unemployed	0.072	0.092
Proportion of 18+ with higher educational qualifications	0.122	0.134
Proportion of employed travelling to work by public transport	0.160	0.157

using the BHPS. The theory behind the exercise is simple. Any variable that is poorly predicted at the individual level by the constraint variables is unlikely to be well modelled by the method (and, although the converse does not necessarily follow, there would be an expectation that the individual and aggregate level correlations would be related). The results of the individual-level regressions of each variable on the set of constraint variables (with the exception of average age, which is an interval measure and hence not comparable with the others) are reported in Table 15.

The most striking aspect of these results is the similarity of the coefficients. It was to be expected that the values of the correlation coefficients would be low, as the dependent variable is binary. What was less predictable was the lack of any clear pattern to the values. The analysis threw little light on the expected aggregate-level relationships.

Table 16 shows the actual (census) and simulated values of the above variables for York.

As can be seen, the method underestimates the York unemployment rate, the percentage of the working population who travel to work by public transport. This can be explained partly by the general difference between the census national rates and the BHPS rates for these variables, which were described in Table 16 and may be due to slight differences in the definitions of these variables.

Figures 23–27 show the scatterplot for each pair of variables at the ward level, the census proportion on the vertical and the simulated proportion on the horizontal axis. A perfect match would find all points on a straight line of gradient 1. Three statistics help to measure the degree of departure from that pattern. As can be seen in Figure 23, there is a relatively good match of simulated and actual values of average age of residents across the 15 wards of York.

Table 15 Coefficients of determination for multiple regressions between SAS constraint variables and BHPS variables, which were not used as constraints in the simulation

Variable	Multiple r
Proportion of adults with long-term illness	0.252
Proportion of economically active unemployed	0.230
Proportion of 18+ with higher educational qualifications	0.233
Proportion of employed travelling to work by public transport	0.250

Table 16 Simulated vs. actual census values for selected variables in York

Variable	Simulation of York 1991	York (census 1991)
Average age of residents	38.2	38.6
Proportion of adults with long-term illness (%)	12.1	12.8
Proportion of economically active unemployed (%)	4.6	7.6
Proportion of 18+ with higher educational qualifications (%)	6.9	8.5
Proportion of employed travelling to work by public transport (%)	4.8	8.0

Nevertheless, as Figure 24 demonstrates, there is a relatively poor match for the values of actual and simulated unemployment rate. Moreover, Figure 25 suggests a further departure from straight line for the values of travel to work by public transport. Likewise, there is a relatively bad match of simulated and actual values of the percentage of adults with higher education as can be seen in Figure 26. Finally, there is slightly better match of the simulated and actual ward values of the percentage of individuals reporting limiting long-term illness (Figure 27). Nevertheless, it should be noted that, at the level of York City, the fit is significantly better.

It can be argued that the difference in the simulated and actual rates of unemployment may be due to a certain extent to different definitions of unemployment adopted by the census and the BHPS. Further, the general underestimation of the percentage of working population who travel to work by public transport may be due to the fact that urban localities were simulated, where the likelihood

Figure 23 Simulated vs. actual average age of residents

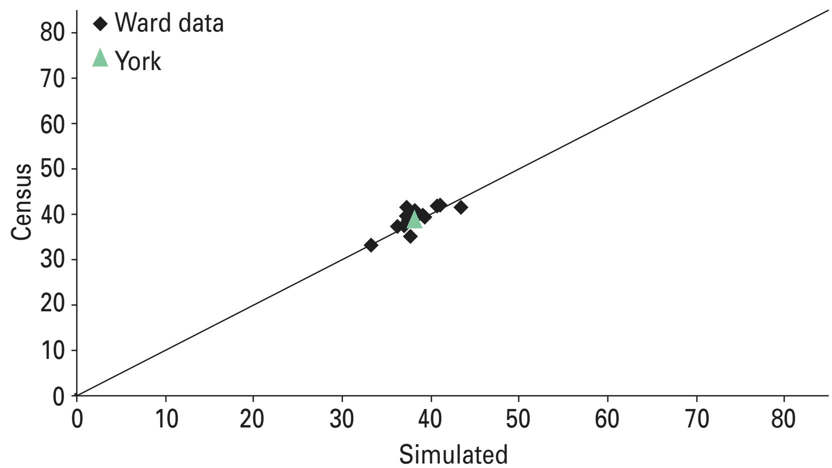


Figure 24 Simulated vs. actual unemployment rate of residents

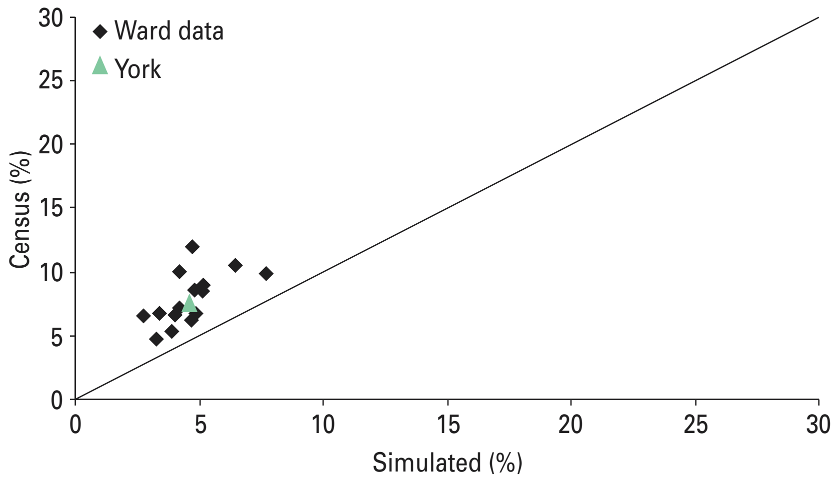


Figure 25 Simulated vs. actual rate of working population travelling to work by public transport

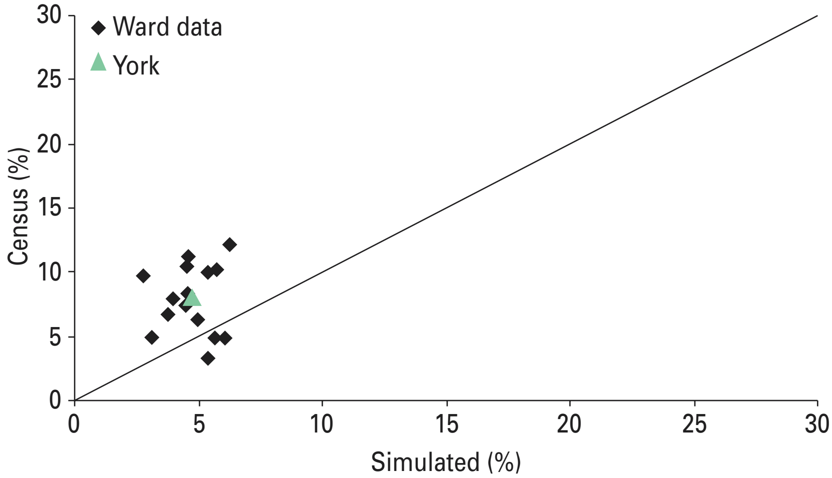


Figure 26 Simulated vs. actual higher degree rate of residents

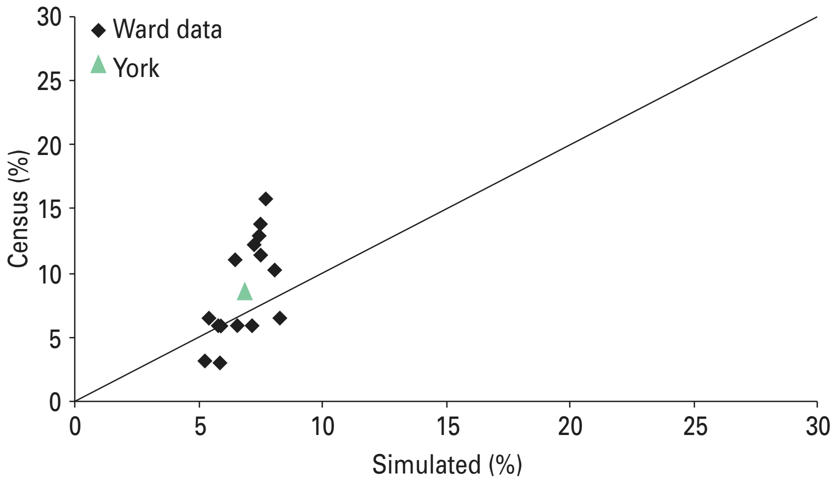
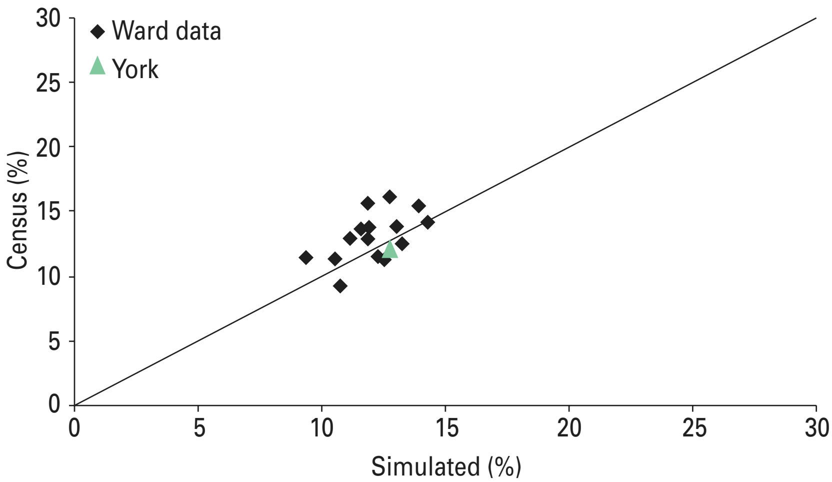


Figure 27 Simulated vs. actual rates of limiting long-term illness



of public transport use is higher. Additionally, the large differences between actual and simulated rates of individuals with higher educational qualifications may be due to the fact that York is a university town.

It would be reasonable to expect that the performance of the model would vary from variable to variable, especially at areas as small as wards and for variables that were not included as constraints in the simulation exercise. Nevertheless, it is encouraging that, as seen in Table 16, the estimates for the City of York match reasonably well their census actual counterparts. It can therefore be argued that the method is quite reliable when analysing socio-economic patterns at the level of the city. This is clearly demonstrated in Figures 28, 29 and 30, which depict the simulated versus the actual rates of limiting long-term illness (LLTI) rates, educational qualification and travel to work by public transport rates respectively for the 641 British parliamentary constituencies.

As can be seen, the fit of the simulated to actual rates at the parliamentary constituency level is significantly better than the respective fit at the small area level (although the simulated data still attenuates). At the ward level, the performance of the model varies considerably and there is a need to introduce further constraints in order to perform analysis at the ward or sub-ward level for particular variables.

Figure 28 Long-term illness (constituencies), *SimBritain* (cill = census data; sill = simulated data)

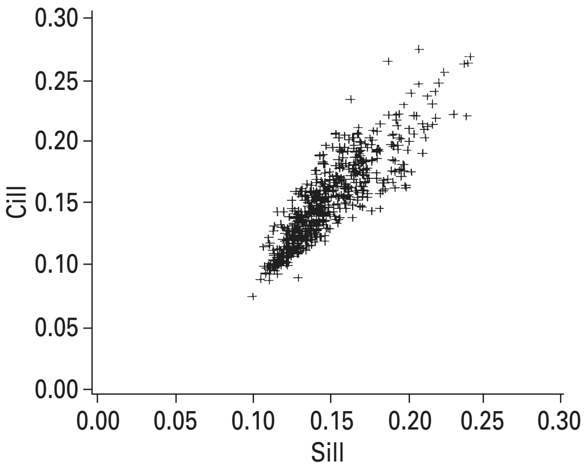


Figure 29 Higher degree rate of residence (constituencies), *SimBritain* (cedu = census data; sedu = simulated data)

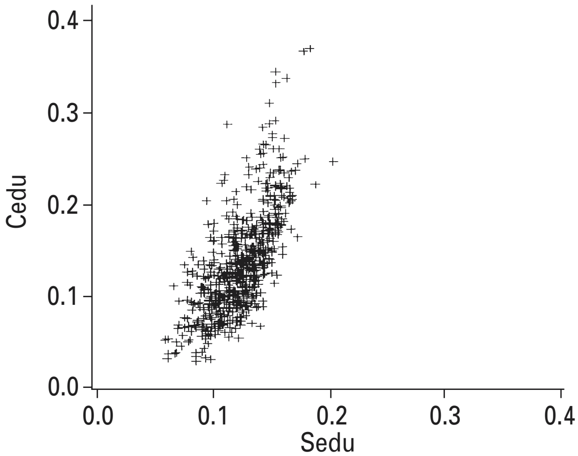
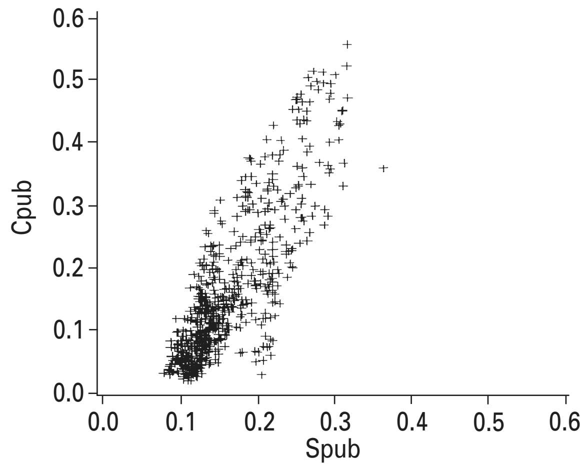


Figure 30 Travel to work by public transport (constituencies), *SimBritain* (cpub= census data; spub = simulated data)



In addition, it seems that the method is not very suitable for the prediction of variables that are affected considerably by external and localised factors, such as transport networks and public transport services, or the presence of a disproportionately large university or a single major employer in the region. This is further demonstrated by Figure 31, which shows the estimated rates of individuals with degrees in Welsh districts. As can be seen, there is a relatively high underestimate of individuals with a university degree in three districts. These are Cardiff, Glamorgan and Ceredigion, which are characterised by the presence of large universities.

Validating the small area population projections

The small area projection method presented in Chapter 10 can be validated by comparing the projection distributions with official government projections. For instance, Figure 32 shows that the national projections made from the 1971, 1981 and 1991 censuses for three categories of car ownership are relatively accurate. The data against which the projections are compared are taken from the General Household Survey (GHS). By 1999, there is some divergence between the projections and the GHS data, with the GHS having a higher proportion of households with one car, but a lower proportion of households with two or more cars. However, it should be noted that there are probably differences in the definitions used for car ownership in the GHS and the census. The census asks about car availability, whereas, in the GHS, the measurement is households with regular use of a car. This difference in definitions could account for the differences between the proportions from the GHS and the proportions in the projections. For example, in Figure 32, it can be seen that there are differences between the census data and the data from the GHS in 1991; if the definitions were the same the proportions should be the same in this year.

Similar comparisons of the simulated trends in other variables (e.g. household types, tenure, etc.) were carried out and were equally satisfactory (note that 2001 census data for the

Figure 31 Simulated vs. actual higher degree rate of residents in Welsh districts, 1991

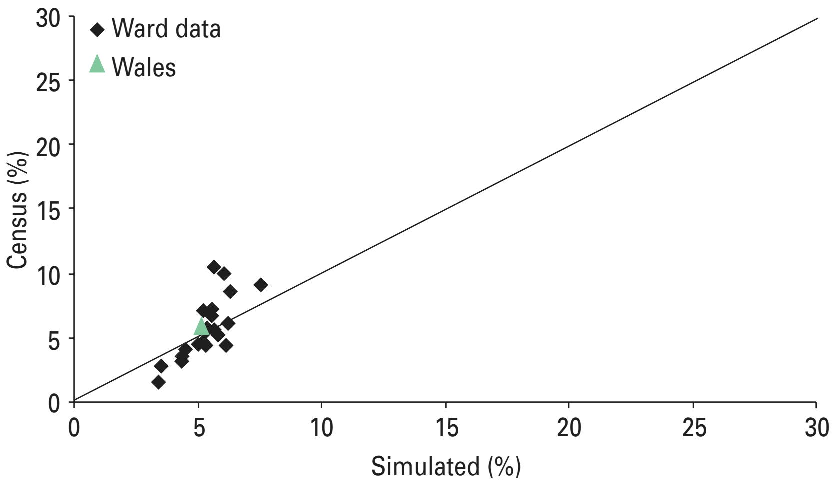
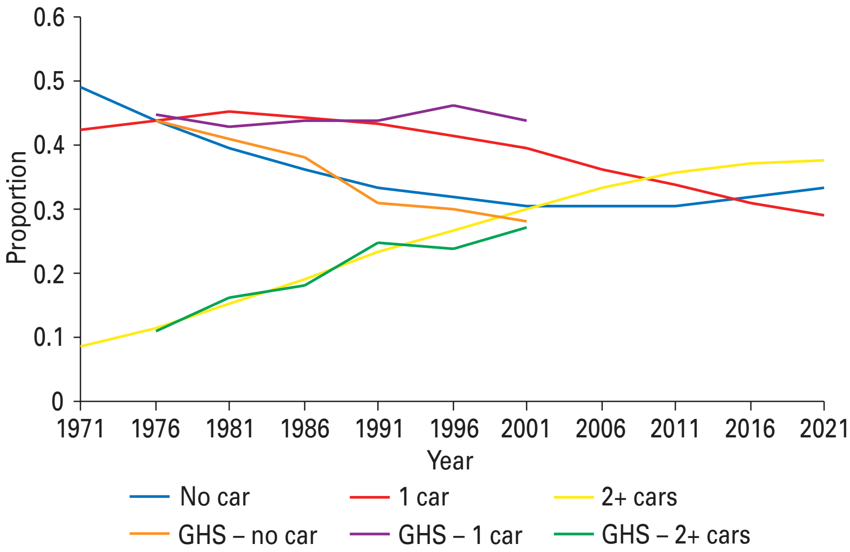


Figure 32 Car ownership in Great Britain, 1971–2021



1991 boundaries of York were not available when the research presented here was conducted).

Another way of checking the reliability of the projection methodology is by using past census data to project distributions of populations into 1991 and then compare the projected values with the actual data from the 1991 census. Table 17 shows an example of a comparison of census data on social class groupings and projected proportions of these groups in 1991. As can be seen, by using the data on social class for the years 1961–71–81 the projection method predicts that 34 per cent of the households in York in 1991 would belong to Class I and II. This prediction matches the actual proportion, which was calculated with the use of 1991 census data. Likewise, the projection method works very well in estimating the 1991 distributions of Class III, IV and V households.

Further, it is possible to validate the projection method by comparing the projected 2001 distribution to the results of the 2001 census returns. Figure 33 shows a comparison of the projected and actual numbers of households in 2001 for Welsh local authorities. Figure 34 depicts a comparison of projected and actual 2001 proportions of households with two or more cars in Welsh local authorities.

Table 17 Comparing census data to projected data for 1991 (projection based on data from the censuses of 1961, 1971 and 1981)

Social class groupings	Census data (%)					Predicted proportion for 1991 (%)	Difference between projection and actual data (%)
	1951	1961	1971	1981	1991		
Class I and II	19	21	24	28	34	34	0
Class III	51	50	49	47	43	44	1
Class IV and V	30	29	27	25	24	22	-2

Figure 33 Projected against actual number of households in each local authority in Wales in 2001

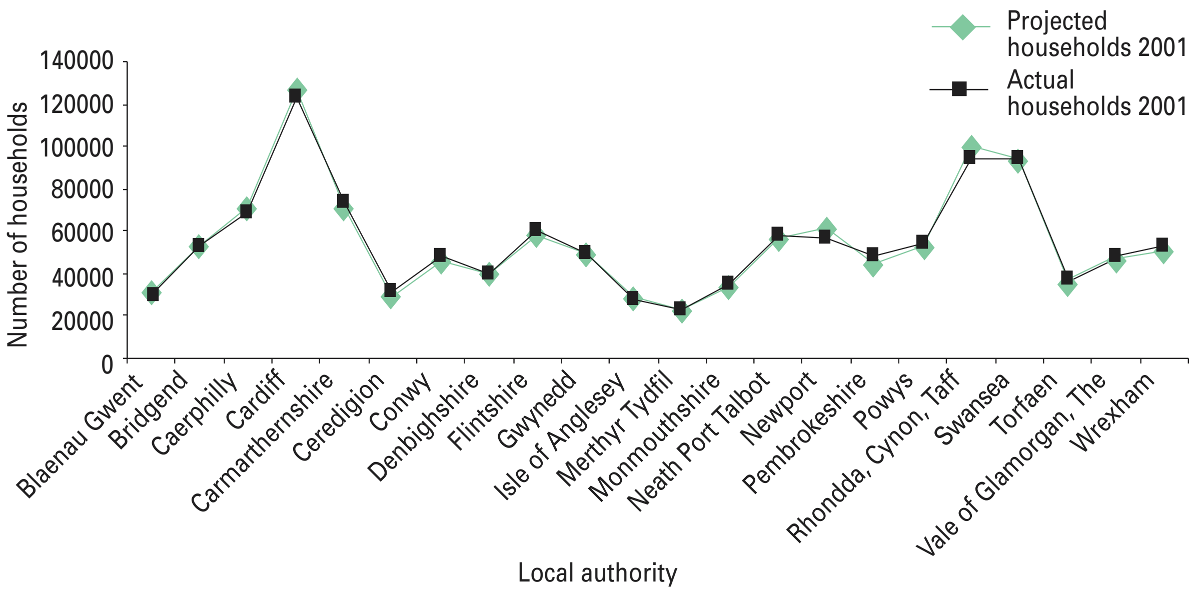
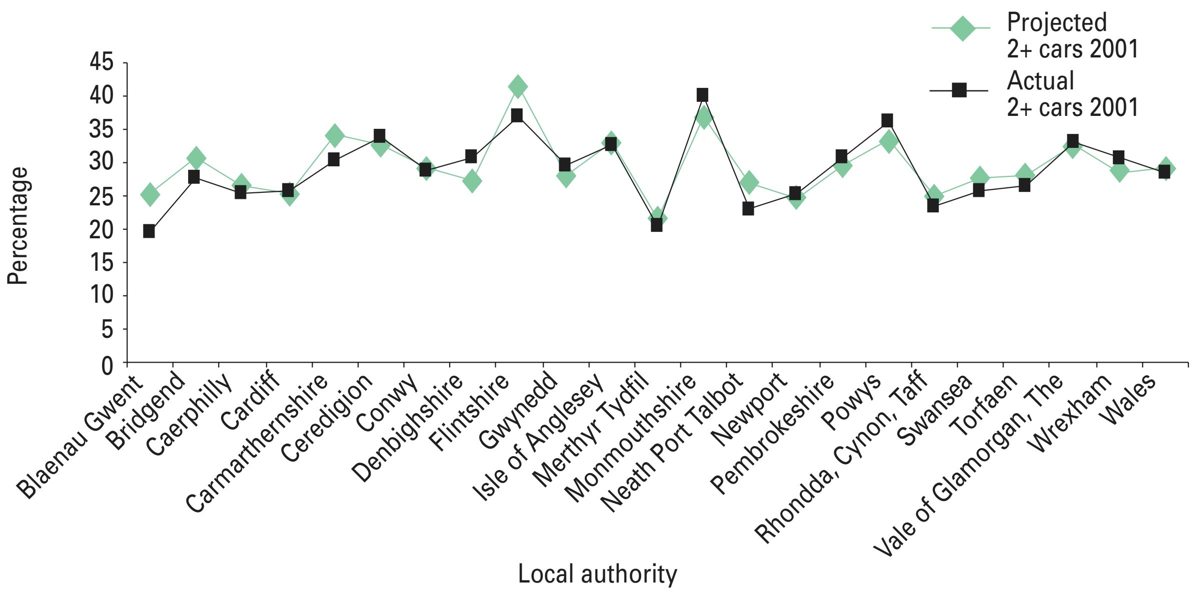


Figure 34 Projected against actual proportion of households with two or more cars in 2001



14 Examples of how policy analysis can be performed

One of the most interesting aspects of the method described in this book is the ability to use the model to perform *what-if policy analysis*. This chapter describes how the outputs of the method can be used to provide policy-relevant data. All households are classified on the basis of their income. Further, the living standards of the simulated households are presented. Some simple evaluations of policies that were introduced by the Labour Government in the late 1990s are examined. In particular, estimates of the impact of the following policies and welfare reforms are discussed:

- Working Families' Tax Credits
- Minimum Wage
- Minimum Income Guarantee
- Winter Fuel Payment
- free TV licence for households where at least one resident is aged 75 or over
- Child Tax Credit and Working Tax Credit (introduced April 2003).

Classifying the simulated households

Before exploring the impact of the above policies on different types of households, all the simulated households were classified on the basis of their income.

In particular the equivalised income (see Box 9 later in this section) estimates for households were used in order to perform an analysis of living standards. The simulated households were divided into the following five groups.

- *Very poor*, comprising all households with equivalised income below or equal to half of the median household income.
- *Poor*, comprising all households with equivalised income greater than half the median and smaller than or equal to three-quarters of the median.
- *Below-average*, comprising all households with equivalised income greater than three-quarters of the median and smaller than or equal to the median.
- *Above-average*, comprising all households with equivalised income greater than the median and smaller than or equal to the median plus a quarter of the median.
- *Affluent*, comprising all households with equivalised income greater than the median plus a quarter of the median.

For the purposes of this book, an example of applying the above classification on the simulated households in York is presented (building on the discussion presented in Chapter 11). Table 18 shows the absolute and relative sizes of each household class throughout the simulation period.

It should be noted that the above classification encapsulates an implicit definition of poverty, by describing the lower income households as *poor* and *very poor*. This is a definition of *relative* poverty, as it is not directly based on the degree to which households are able to satisfy their physiological or other basic needs. However, it should be noted that the concept of poverty constantly evolves and it has long been argued that the subsistence approach to the definition of poverty is inadequate. As Gordon and Pantazis (1997) point out:

> *The subsistence approach to the definition of poverty is an 'absolute' concept of poverty; it is dominated by the individual's requirements for physiological efficiency. However, this is a very limited conception of human needs, especially when considering the roles men and women play in society. People are not just physical beings, they are social beings. They have obligations as workers, parents, neighbours, friends and citizens that they are expected to meet and which they themselves want to meet.*
> (Gordon and Pantazis, 1997, p. 9)

It can be argued that equivalised household income is probably the most appropriate variable that can be used for the analysis of poverty, as it reflects the degree to which people are able to satisfy their basic needs as well as fulfil their various roles in society. Further, given that the analysis presented here projects the population of York into the future, it can be argued that income should be used to define and analyse poverty, as it will be likely to keep its significance through time (whereas human needs and social roles will evolve over time).

Table 18 The size of the simulated household classes, 1991–2021

Class size by year	Very poor	Poor	Below average	Above average	Affluent	Total number of households
Absolute size						
1991	7,190	7,149	6,589	5,322	15,605	41,855
2001	8,208	9,373	6,020	6,753	16,848	47,202
2011	9,085	9,149	7,303	8,293	17,244	51,074
2021	11,700	6,222	9,476	11,185	16,213	54,796
% of all households						
1991	17.2	17.1	15.7	12.7	37.3	100.0
2001	17.3	19.9	12.8	14.3	35.7	100.0
2011	17.8	17.9	14.3	16.2	33.8	100.0
2021	21.3	11.4	17.3	20.4	29.6	100.0

Box 9 Equivalence scales

It should be noted that one of the problems with using household income as a classification variable is that different types of households have differing needs, which arise from differences in size, age of household members, etc. As Atkinson (1983) points out, the simplest way of allowing for differing household needs is to treat all household members as having the same needs and to calculate the household income per head. Nevertheless, this simple way of allowing for variation in household needs fails to recognise the role of age and the possible economies of scale that result from cohabitation.[1] Atkinson (1983) reviews some of these attempts and provides a useful comparison, which is reproduced in Table 19.

(Continued)

Table 19 Comparing equivalence scales

Household type	Rowntree	Supplementary benefits (UK)	US poverty scale	McClemence (UK)	Lazear and Micahael
Single person	60	62	80	61	94
Couple	100	100	100	100	100
And 1 child	124	121	119	121	122
And 2 children	161	132	152	142	139
And 4 children	222	184	201	184	––

Source: Atkinson, 1983, p. 49.

In the context of the work reported here, the *McClemence Equivalence Scales* were used to adjust household income, so that meaningful comparisons can be made. The calculated *McClemence Equivalence indices* for each household are included in the BHPS. In particular, the *McClemence Equivalence Scale* is the semi-official scale for the UK and has also been used in publications such as the Households Below Average Incomes (HBAI). The *McClemence Equivalence Scale* is described in some detail in Table 20.

Table 20 The McClemence Equivalence Scale

	Before housing costs	After housing costs
Head	0.61	0.55
Spouse	0.39	0.45
Other second adult	0.46	0.45
Third adult	0.42	0.45
Further adult	0.36	0.40
Dependent child aged:		
0–1	0.09	0.07
2–4	0.18	0.18
5–7	0.21	0.21
8–10	0.23	0.23
11–12	0.25	0.26
13–15	0.27	0.28
16+	0.36	0.38

Source: Taylor *et al.*, 2001.

Exploring the living standards of different household classes

One of the advantages of spatial microsimulation is that it enables researchers to investigate the living standards of individuals and households using the simulation outputs. For instance, it is possible to draw pictures of the life of households in a similar way as Rowntree (2000) did in his original study of poverty in York, where he provided extracts of notes on the characteristics and lives of households placed between various classes.

Table 21 gives an example of how this analysis can be replicated by making notes on a selection of households. In particular, as pointed out in Chapter 9, it is possible to use the output of the method presented here to make some notes on the life of simulated households, and in this case of typical very poor households.

Table 22 lists some statistics on the living standards of these households for the period 1991–2021.[2] As can be seen, the poorest segment of the York society is predicted as a group to increase in size, from 17.2 per cent of total

Table 21 Example of notes on simulated *very poor* households

Age of household head(s)	Description
72	Single-person household, female, retired assembler/lineworker (electrical/ electronic goods). Feels that is just about getting by financially. House owned outright. Believes that all health care should be free.
56 and 52	Married couple, male aged 56, economically active but unemployed. Formerly employed as motor mechanic/auto engineer. Female aged 52, economically inactive. Food expenditure: £25 per week. No car. House owned with mortgage. Highest educational qualification of male: GCE O levels. Female has no formal qualifications.
46	Female, divorced, full-time personal services worker (hairdresser on seasonal/ temporary job). Finding it quite difficult financially. No dependent children. Believes that all health care should be free, feeling unhappy or depressed. Has one car. Weekly household food expenditure: £35.
78	Divorced, female, retired sales assistant. Feels that there is no one who she could count on to listen if she needs to talk. Feels that is just about getting by financially. No children, no car. Weekly food and grocery expenditure: £20.
34 and 21	Married couple, two children (aged eight and five). Two cars. Male has full-time job. He is self-employed (craft and related occupations – construction). Female is economically inactive (family care). Just about getting by financially. Household weekly food and grocery expenditure: £40.
34 and 33	Married couple, Four children (aged 14, 13, 12 and 9). Male unemployed (previous job: food, drink and tobacco process operative) Female economically inactive (family care). Household weekly expenditure on food: £80.
30	Female divorced, mother of three children (aged 11, 9 and 2), economically inactive (family care), no qualifications, formerly employed as a sales assistant. No car. House rented from local authority. Weekly expenditure on food: £20.

Table 22 Living standards of *very poor* households

Very poor households	1991	2001	2011	2021
Households (% of all households in York)	17.2	17.3	17.8	21.3
Individuals (% of all individuals in York)	14.7	13.3	13.7	20.5
Children (% of all children in York)	21.8	17.7	18.6	38.5
LLTI (as a % of all individuals in group)	9.0	7.3	5.4	7.9
Older people (over 64 years as a % of all individuals in group)	30.1	32.0	33.3	44.2
Individuals in group with father's occupation: unskilled (%)	10.5	6.8	3.3	15.1
Reporting anxiety and depression (% of all individuals in group)	10.6	10.3	7.4	3.1
Individuals who reported that they have no one to talk to (%)	19.9	23.8	31.1	31.5
Promotion opportunities in current job (as % of individuals with a job)	33.7	36.9	51.9	79.7
Feeling unhappy or depressed (%)	19.9	19.0	18.2	12.1
Home computer in accommodation (%)	1.4	1.0	0.5	0.4
House without central heating (%)	26.1	21.4	21.4	31.1
Single-person households (%)	61.6	76.0	77.9	64.4
Cars–households ratio	0.23	0.32	0.38	0.40

households in 1991 to 21.3 per cent in 2021. Further, the number of children living in *very poor* households rises significantly from 21.8 per cent in 1991 (as a percentage of all children in York) to 38.5 per cent in 2021. Likewise, the number of older people in this group increases from 30.1 per cent in 1991 to 44.2 per cent in 2021. The incidence of Limiting Long Term Illness (LLTI) is estimated to be 9 per cent in 1991 and is predicted to fall to 7.9 per cent by 2021. Further, an estimated 10.6 per cent of the population in 1991 report anxiety and depression problems.

It is interesting to note that, in 1991, it was estimated that 10.5 per cent of very poor households had a household head whose father had an unskilled occupation. This percentage is projected to rise in 2021 to 15.1 per cent.

Moreover, 19.9 per cent of individuals in the *very poor* group reported that they felt they could not count on anyone to listen to them if they needed to talk. It can be argued that this proportion can be seen as an index of 'loneliness' and it is noteworthy that it is projected to increase significantly in 2021, when 31.5 per cent of the individuals in the very poor household group are estimated to have increased feelings of 'loneliness'.

A useful indicator of well-being and prosperity is the ratio of cars–households, especially given that there is a general increasing trend in car ownership across all households in the simulation period. Nevertheless, there are only slight increases in this ratio in the *very poor* households, in the period 1991–2011. The ratio increases from 0.23 in 1991 to 0.40 in 2021. As will be seen later, in affluent households, this variable increases from 0.94 to 1.72. Further, the percentage of households that have a home computer is estimated to be 1.4 per cent in 1991 and it is projected to drop to 0.4 per cent in 2021. It should be noted, though, that this projection is not very realistic, given that home computers become increasingly common in households. However, the home computer here may be seen as the equivalent of a high-tech product at any time (e.g. in 2001 it could be a DVD player or mobile phone with photo-messaging and in 2021 it may be virtual reality facilities or some other product or service).

Moreover, it is worth noting that only 33.7 per cent of individuals who had a job felt that they had opportunities for promotion in 1991. This percentage increases to 79.7 per cent by 2021.

In contrast to the projected change in size of the *very poor* household class, there is a decreasing trend in the numbers of households belonging to the slightly better off class, which was described above as *poor households*. Table 23 includes some notes on typical households of this type.

Table 24 gives some statistics on the living standards of these households, which comprised 17.1 per cent of the total households in York in 1991. This percentage drops to 11.4 per cent in 2021. It is interesting to note that the number of children in this group increases throughout the simulation period. In particular, the percentage of children in *poor* households is simulated to be 15.4 per cent in 2021, whereas, in 1991, it was estimated at 12.7 per cent. In contrast, the proportion of older people in *poor* households falls from 30.6 per cent in 1991 to 17.3 per cent in 2021. It should be noted, however, that these projections did not take into account welfare policy initiatives, introduced in the late 1990s, such as the Working Families' Tax Credit.

It is interesting to note that individuals in the households of this group have much higher rates of limiting long-term illness (LLTI) than their

Table 23 Typical *poor households*

Age of household head(s)	Description
32 and 28	Married couple, male works in construction trade (builder), female works in health and related occupations (care assistant), two dependent children aged 3 and 2, weekly food: £40, one car, house owned with mortgage.
71 and 70	Local authority renting, married couple, no children, male retired, female – family care, £35 on food weekly.
63 and 59	Married couple, male retired sales assistant, female long-term sick/disabled, last job manager in service industry, no children, house owned outright, weekly food expenditure: £50, doing all right financially.
29 and 25	Two male agricultural workers, never married, one self-employed agricultural machinery (CSE Grade 2) driver and the other (no qualifications) employed in farming and related occupations, household weekly food expenditure: £70.
47 and 46	Married couple, male has a teaching qualification, employed as a driving instructor, female no qualifications, employed in childcare, weekly expenditure on food: £50, house owned with mortgage.

Table 24 Living standards of *poor* households

Poor households	1991	2001	2011	2021
Households (% of all households in York)	17.1	19.9	17.9	11.4
Individuals (% of all individuals in York)	14.8	17.8	17.8	11.7
Children (% of all children in York)	12.7	19.2	25.3	15.4
LLTI (as a % of all individuals in group)	14.1	17.5	14.3	7.1
Older people (over 64 years as a % of all individuals in group)	30.6	31.1	28.5	17.3
Individuals in group with father's occupation: unskilled (%)	6.1	6.0	5.6	1.2
Reporting anxiety and depression (% of all individuals in group)	9.2	4.9	3.5	6.6
Individuals who reported that they have no one to talk to (%)	11.6	7.0	5.1	10.0
Promotion opportunities in current job (as % of individuals with a job)	32.8	39.2	34.4	15.0
Feeling unhappy or depressed (%)	23.8	18.5	12.2	16.4
Home computer in accommodation (%)	2.1	5.1	7.6	11.7
House without central heating (%)	42.8	39.3	45.4	63.5
Single-person households (%)	48.6	60.4	60.5	63.7
Cars–households ratio	0.47	0.58	0.81	0.68

counterparts in *poor* households. In particular, 14.1 per cent of the population in this group report LLTI in 1991. Nevertheless, this rate drops to 7.1 per cent in 2021. There are generally high rates of individuals in *poor* households, feeling unhappy or depressed in that group throughout the simulation period. Nevertheless, unlike the *very poor* households, there are generally smaller proportions of individuals who feel that they have no one to listen to them if they need to talk. These smaller proportions may be partially explained by the relative stability of the percentage of single-person households.

It is noteworthy that there is a decrease in the proportion of households that come from unskilled backgrounds. This proportion is estimated to be 6.1 per cent in 1991 and is projected to fall to 1.2 per cent in 2021. There is also a significant decrease in the proportion of individuals with a job who feel that they have promotion opportunities. Also, the car to household ratio and the percentage of households that have a home computer increase throughout the 30-year simulation.

It is interesting to explore the possible causes of low income in the simulated database. Tables 25 and 26 list some of the potential variables that may be associated with low incomes and poverty.

As can be seen, almost half (45.4 per cent) of the economically active individuals living in *very poor* households are unemployed in 1991. It therefore seems that, although unemployment remains an important determinant of poverty, there are other factors that contribute significantly to poverty, as the simulation predicts near-full employment conditions in the future.

Table 25 ***Very poor*** **households, possible causes of poverty**

Very poor households	1991	2001	2011	2021
Unemployed (as a % of economically active in group)	45.4	25.7	16.7	9.6
Economically active (%)	18.3	17.1	16.8	17.7
Vocational qualifications (% of all adult individuals in group)	20.9	20.7	18.9	12.2
Full-time job (% of economically active in group)	43.1	65.9	80.7	90.1
Adults with no qualifications (% of all adult individuals in group)	58.4	65.2	72.3	78.9

Table 26 ***Poor*** **households, possible causes of poverty**

Poor households	1991	2001	2011	2021
Unemployed (as a % of economically active in group)	12.7	3.5	0.0	2.0
Economically active (%)	14.1	17.5	14.3	7.1
Vocational qualifications (% of all adult individuals in group)	23.2	20.3	19.6	16.5
Full-time job (% of economically active in group)	68.1	68.9	77.8	76.2
Adults with no qualifications (% of all adult individuals in group)	58.2	64.0	63.4	60.4

Table 25 shows that there is an increasing trend in the proportion of individuals without any qualifications living in *very poor* households. Also, there is a decreasing trend in the numbers of individuals with vocational qualifications. It can be argued that the lack of educational qualifications may be one of the major causes of low pay and limited chances of finding a secure well-paid job. It should be noted, though, that, given the increasing trend in general education levels, the *no-qualifications* variable in the future may mean limited qualifications, rather than no qualifications at all.

It is worth exploring further the distribution of individuals without qualifications across different classes. The age and gender-specific *no-qualifications* ratio was calculated for the whole York population and then applied to the totals of each household class, in order to estimate the expected numbers of individuals without qualifications by class. The observed over-expected ratio was then calculated for all individuals over 15 years old. Table 27 summarises these ratios for all household classes throughout the simulation period.

The last column of Table 27 shows the ratio between the values for very poor and affluent. As can be seen, this is estimated to be *2.7* in 1991, which means that individuals belonging to the very poor class were *2.7* times more likely to have no qualifications than their counterparts in the affluent class having allowed for their age distribution. Further, this ratio is projected to rise to *4.5* by 2021.

The statistics presented here can be extremely useful in the design of anti-poverty policies. It can be argued that policies aiming at the enhancement of educational qualifications and the increase of employment opportunities would lead to a significant improvement of the living conditions of the groups of households explored here. Nevertheless, any policy that

Table 27 Observed over-expected ratio for individuals over 15 years old without any qualifications

Year/group	Very poor	Poor	Under average	Over average	Affluent	*Very poor/ affluent*
1991	1.73	1.36	1.08	0.69	0.63	*2.7*
2001	1.51	1.40	0.7	0.77	0.58	*2.6*
2011	1.36	1.29	0.75	1.05	0.45	*3.0*
2021	1.54	1.08	0.57	0.95	0.34	*4.5*

Figure 35 *Very poor* households, sources of income 1991–2021

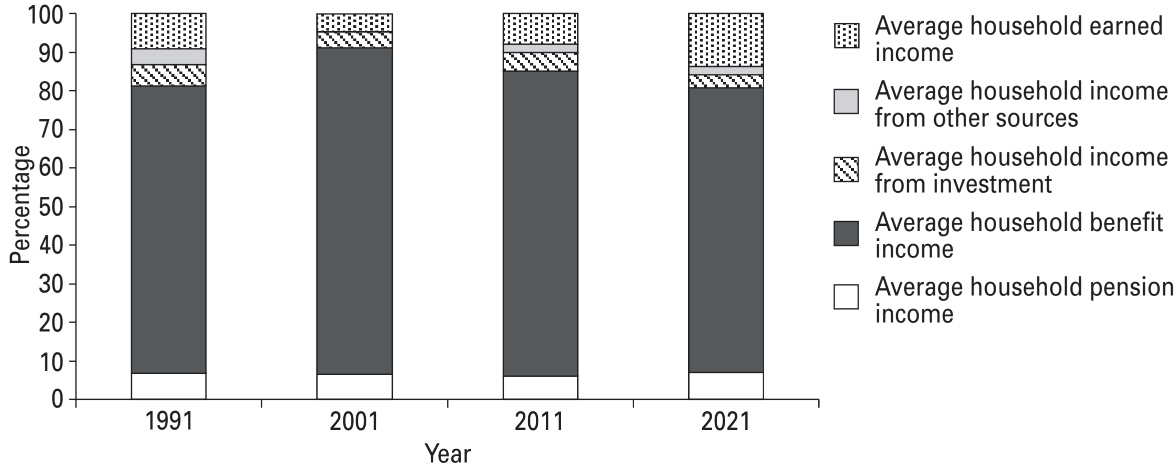
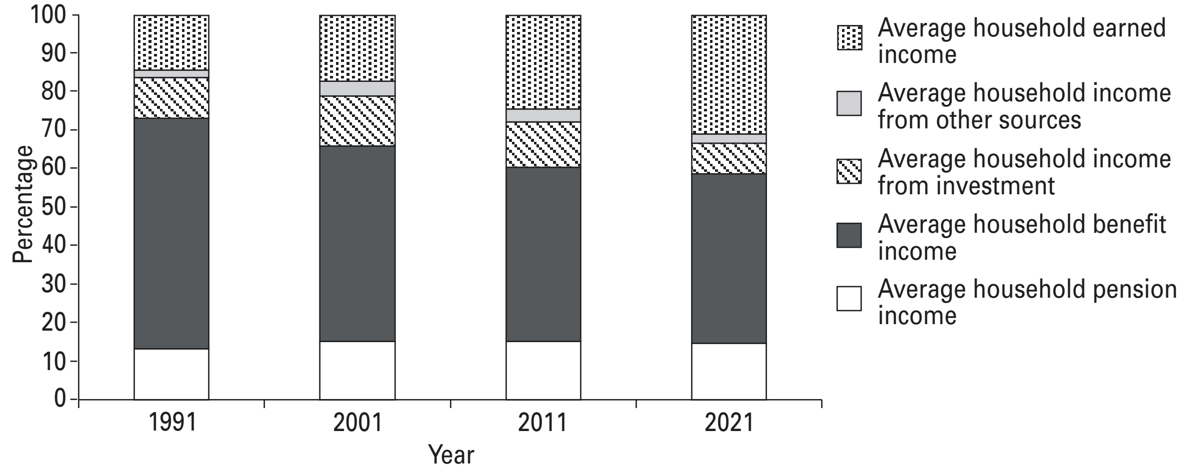


Figure 36 *Poor* households, sources of income 1991–2021



would aim at addressing the position of individuals living in *poor* households would not produce results in the short term. Therefore, it may be useful to enhance such policy measures by increasing the sources of income received by these households. Figures 35 and 36 show the proportion of average income from different sources for *very poor* and *poor* households. As can be seen, benefit income is the major component of the total household income of the *very poor* households, throughout the simulation period. In contrast, the levels of earned income are quite low and have decreasing trends.

On the other hand, poor households have a larger proportion of their income from labour.

It is possible to repeat the above analysis for the other household classes. The following section shows how it is possible to evaluate the impact of various policies on different household types and different household classes.

Estimating policy impacts

The above discussion and classification of households is based on 1991 household incomes. Therefore, in order to enable the impact analysis of policy reforms that occurred in the late 1990s, the amounts pertaining to different policies and benefit rules must be adjusted to allow for inflation. For instance, the Minimum Wage in October 2002 was £4.50 per hour for individuals at work who are over 21 years old and £3.80 for individuals aged 18–21.[3] Clearly the equivalent of these figures in 1991 would have been lower and there is a need to readjust them accordingly (see Box 10).

Box 10 Readjusting monetary amounts
One way of readjusting the policy-related amount is by using the growth of the Retail Price Index (RPI) from 1991 until now. The RPI is the UK's principal measure of consumer price inflation and is defined as 'an average measure of change in the prices of goods and services bought for the purpose of consumption by the vast majority of households in the UK' (ONS, 2003d). Table 28 depicts the growth of RPI between 1991 and 2002.

Table 28 The growth of RPI between 1991 and 2002

Year	RPI
1986	100
1987	101.9
1988	106.6
1989	113.1
1990	121.4
1991	129.5
1992	135.1
1993	139
1994	141.3
1995	144.5
1996	148.2
1997	151.5
1998	154.5
1999	157.1
2000	159.9
2001	163.7
2002	167.5
Change 1991–2002 (%)	29.3

Retail Price Index (RPI): all items excluding mortgage interest payments and indirect taxes
(RPIY) (January 1987 = 100).

Source: Office for National Statistics (http: // www.statistics.gov.uk /).

In the context of this book, all the figures that related to various policies and benefit rules were readjusted on the basis of RPI growth. These adjusted data were then used to explore the impact of a selection of important policies.

Box 11 A selection of policies to be evaluated

Working Families' Tax Credit

One of the major policy initiatives that was implemented in the 1990s was the Working Families' Tax Credit (WFTC), which is an allowance paid to low-paid workers with children (Fitzpatrick *et al*., 2002; Inland Revenue online,[4] 2003). In order to qualify for WFTC, individuals would have to fulfil the following criteria.

- They or their partner should work normally full time (16 hours or more a week).
- They have at least one dependent child for whom they are responsible.
- They do not receive Disabled Person's Tax Credit.
- Their income is sufficiently low.
- Their savings and capital are not worth more than £8,000.
- They are present and ordinarily resident in Great Britain.
- They are not subject to immigration control.

WFTC is calculated by comparing the family income with the applicable amount or threshold figure, which in 2002 was £94.50. If the family income is less than the applicable amount, then the family receives the maximum WFTC. If the family income exceeds the applicable amount, the maximum WFTC is reduced by 55 per cent of the excess (Fitzpatrick *et al*., 2002). As noted above, in the context of the research reported here, all the relative amounts were adjusted to allow for inflation. In the case of WFTC, the applicable amount of £94.50 in 2002 was readjusted to its equivalent in 1991 on the basis of the RPI growth of 29.3 per cent. Thus, the adjusted applicable amount that we used was £66.77. Further, all the relevant credits were adjusted before allocating them to eligible households of the simulated database. Table 29 lists the actual (2002) and adjusted (1991) amounts for the various credits.

Minimum Wage

Another related major policy development in the 1990s was the introduction of the Minimum Wage. As noted above, the Minimum Wage in October 2002 was £4.50 per hour for individuals at work who were over 21 years old and £3.80 for individuals aged 18–21, which were adjusted to £2.97 and £2.54 for 1991.

(Continued overleaf)

Table 29 Working Families' Tax Credits

Working Families' Tax Credits	Amount in 2002–03[a]	Adjusted for 1991
Couple or lone parent	£60.00	£42.39
Child aged		
Under 16	£26.35	£18.62
16–18	£27.20	£19.22
30 hours' credit	£11.65	£8.23
Disabled Child Credit	£35.50	£25.08
Enhanced Disability Credit		
Couple or lone parent	£16.25	£11.48
Child	£46.75	£33.03
Childcare Credit		
One child	70% of up to £135	70% of up to £95.39
Two or more children	70% of up to £200	70% of up to £141.31
Additional partners in a polygamous marriage	£22.70	£16.04

a Fitzpatrick *et al.* (2002, p. 560)

Minimum Income Guarantee

The introduction of the Minimum Income Guarantee was another major policy development that occurred in the late 1990s. This guarantee aimed at topping up the income of elderly individuals or couples (aged 60 or over and with savings less than £12,000) to a minimum level. This minimum level is currently (March 2003) £98.15 for a single person and £149.80 for a couple. These figures were adjusted on the basis of RPI growth to £69.35 and £105.84.

Winter Fuel Payment and free TV licence for older people

Another policy initiative that was aimed at boosting the incomes of older people was the Winter Fuel Payment, which is given to individuals aged 60 or over. This amount was £200 in 2003 and was adjusted to £141.31 for 1991. Further, a similar government initiative was the provision of free TV licences to all individuals aged 75 and over. In the case of TV licences, there is no need to readjust the 2002–03 figure to 1991, as data exist on the TV licence across time. The TV licence was currently £112 in 2002, whereas, in 1991, it was £77.[5]

Once all the figures were adjusted, the next step was to estimate the redistributive effects that these policies would have if they had been implemented in each of the simulation years. Table 30 summarises the estimated increase that would occur to the average incomes of households by class, as defined in the previous section.

As can be seen, the policy changes would contribute to the significant increase of the average income of the *very poor* households. In particular, according to the simulation outputs, the income of these households would on average increase by 17.9 per cent in 1991. Further, the income of *poor* and *below average* households would also increase by 7.14 per cent and 4.18 per cent respectively. On the other hand, there would be very small proportional increases in the incomes of *above average* and *affluent* households. Similar redistributive

Table 30 Simulated impact of policy changes by household class and simulation year

	Extra income (£) (in 1991 terms)	Extra income (£) (in 2003 terms)[a]	Income increase (%)	Income increase as % of all income in York
1991				
Very poor	5,445,873.98	7,041,515.06	17.9	0.998
Poor	3,627,500.42	4,690,358.04	7.14	0.679
Below average	2,625,234.42	3,394,428.11	4.18	0.481
Above average	1,787,499.24	2,311,236.52	2.61	0.327
Affluent	3,232,856.34	4,180,083.25	0.36	0.592
2001				
Very poor	6,480,555.59	8,379,358.38	17.8	0.996
Poor	4,477,920.75	5,789,951.53	6.41	0.707
Below average	3,097,989.51	4,005,700.44	4.92	0.476
Above average	3,756,751.62	4,857,479.84	4.11	0.577
Affluent	527,218.68	681,693.75	0.14	0.081
2011				
Very poor	8,346,653.06	10,792,222.41	19.5	1.091
Poor	4,935,563.72	6,381,683.89	6.72	0.676
Below average	6,824,794.89	8,824,459.79	9.30	0.892
Above average	3,432,418.08	4,438,116.58	2.85	0.449
Affluent	105,247.32	136,084.78	0.02	0.014
2021				
Very poor	13,130,269.94	16,977,439.04	21.73	1.591
Poor	3,574,430.01	4,621,738.00	6.87	0.433
Below average	9,981,672.96	12,906,303.14	4.40	1.210
Above average	4,305,714.18	5,567,288.43	2.45	0.522
Affluent	7,754.22	10,026.21	0.00	0.001

a Assuming that the growth of income for all household groups was equivalent to the RPI growth for the period 1991–2003.

patterns are observed in all simulation years. It should be noted that, although the increases in the average incomes of *very poor* households tend to be relatively high, these are very small when seen as a proportion of the sum of all household incomes in York (see last column of Table 30). Nevertheless, it should be noted that the above estimates of income increases for different household classes are based on the assumption of full benefit take-up. Further, any previous benefits that were substituted by the new policy measures were not taken out of the original household income.[6]

It is interesting to note that the suggested policies would have a great impact on families with children. For instance, according to the 1991 simulation outputs, there would be 246 children living in families whose income would increase by 54.1 per cent. Further, there would be 486 children living in families that would experience income increases of over 15.4 per cent. Table 31 lists the estimated average increase to the incomes of households with one or more dependent children in all simulation years.

As can be seen, the average increase in the income of households with dependent children rises to 5 per cent in 2021. This is because of the high projected numbers of children living in *poor* and *very poor* households in 2021.

It is interesting to use the BHPS to draw a picture of typical households that would be affected by the policy changes. Table 32 gives a description of typical simulated households that

Table 31 Average increase in the income of households with dependent children (without removing the impact of previous policies such as Family Credit)

Year	Average increase (%)	Increase as % of income in York
1991	2.2	0.47
2001	2.2	0.39
2011	2.8	0.42
2021	5.0	0.66

Table 32 Typical simulated households that would be most affected by the 1990s' welfare reforms

Age of household head(s)	Description
18 and 18	Married couple, one newborn baby. Male no qualifications, working in sales and services. Female GCE O levels, in family care (formerly employed in sales and services). Weekly expenditure on food: £20. Household income before policy effects: £6,265.34. Income after policy effects: £9,656.62 (increase of 54.1 per cent). No car.
26 and 22	Married couple, one child aged 3. £9,230.02; both in full employment, full time. Male plant and machine operative, female sales and services. Male has CSE (Grade 2–5) qualifications. Female has GCE O levels. Average food expenditure per week: £30. One car. Income after policy change: £11,952.44 (increase 29.5 per cent).

would be most affected by the 1990s' welfare reforms.

It is also interesting to note that the model suggests that several households would change class (e.g. from *very poor* to *poor*) under the suggested changes. Table 33 lists the class transitions by year.

As can be seen the largest number of class transitions would occur had the policies been adopted in 1991, when 3,270 households would have moved from *very poor* to *poor*.

Another way of examining the impact of the above policy change is by analysing the effect of these changes on the income distribution across household deciles. It is useful at this stage to utilise research on the income distribution in Britain carried out by the Institute for Fiscal Studies (IFS). Table 34 describes the monthly income levels for different household types (Shephard, 2003), by the income decile they fall in.

It is interesting to examine the numbers of households in York that fall into the different national income distribution deciles. Table 35 shows how many of the simulated households (in 2001) of each type in York fall into the IFS estimated income distribution.[7] Table 36 shows these households as a proportion of all households of each type in York. Further, Tables 37 and 38 show how this distribution would be affected by the policy changes described above.

Table 33 Class transitions triggered by policy changes

Class Transitions	Households	% of all households
In 1991		
From *very poor* to *poor*	3,720	8.89
From *poor* to *below average*	1,137	2.72
From *below average* to *over average*	774	1.85
From *above average* to *affluent*	866	2.07
In 2001		
From *very poor* to *poor*	2,782	5.89
From *poor* to *below average*	770	1.63
From *below average* to *over average*	790	1.67
From *above average* to *affluent*	824	1.75
In 2011		
From *very poor* to *poor*	1,150	2.25
From *poor* to *below average*	617	1.21
From *below average* to *over average*	2,565	5.02
From *above average* to *affluent*	1,652	3.23
In 2021		
From *very poor* to *poor*	2,280	4.16
From *poor* to *below average*	3,238	5.91
From *below average* to *over average*	54	0.10
From *above average* to *affluent*	259	0.47

Table 34 Where do you fit in?

	Monthly income levels (£)		
	Single person, no children	Couple, no children	Couple with two children (aged 4 and 13)
Bottom decile	0–400	0–700	0–1,000
Decile 2	400–500	700–900	1,000–1,200
Decile 3	500–600	900–1,000	1,200–1,500
Decile 4	600–700	1,000–1,200	1,500–1,700
Decile 5	700–800	1,200–1,400	1,700–2,000
Decile 6	800–900	1,400–1,600	2,000–2,300
Decile 7	900–1,100	1,600–1,800	2,300–2,600
Decile 8	1,100–1,300	1,800–2,100	2,600–3,100
Decile 9	1,300–1,700	2,100–2,800	3,100–4,000
Top decile	1,700+	2,800+	4,000+

Note: Incomes are monthly incomes measured before housing costs and are expressed in 2001–02 prices. The income differences across family types reflect the 'equivalence scales' used. Income ranges within each decile group are the same once adjusted for household size and composition.

Source: Shephard's calculations using Family Resources Survey (Shephard, 2003, p. 5).

Table 35 Where do households in York fit in? (households in 2001, before policy changes)

	Number of household type		
Decile	Single person	Couple with no children	Couple with two children
Bottom decile	6,236	477	200
Decile 2	2,825	616	397
Decile 3	3,277	236	197
Decile 4	330	1,142	237
Decile 5	756	1,181	449
Decile 6	561	1,027	495
Decile 7	1,906	1,076	349
Decile 8	2501	982	382
Decile 9	416	1,350	171
Top decile	1,485	2,864	0
Total	20,293	10,951	2,877

Table 36 Households type by decile as a proportion of all households of this type (2001, before policy changes)

Decile	Proportion of household type (%) Single person	Couple with no children	Couple with two children
Bottom decile	30.7	4.4	7.0
Decile 2	13.9	5.6	13.8
Decile 3	16.1	2.2	6.8
Decile 4	1.6	10.4	8.2
Decile 5	3.7	10.8	15.6
Decile 6	2.8	9.4	17.2
Decile 7	9.4	9.8	12.1
Decile 8	12.3	9.0	13.3
Decile 9	2.0	12.3	5.9
Top decile	7.3	26.2	0.0

Table 37 Where do households in York fit in? (transitions in 2001, after policy changes)

Decile	Number of household type Single person	Couple with no children	Couple with two children
Bottom decile	-5,097	0	0
Decile 2	4,496	-170	0
Decile 3	316	102	0
Decile 4	285	10	0
Decile 5	0	-539	0
Decile 6	0	597	-160
Decile 7	0	0	160
Decile 8	0	0	0
Decile 9	0	0	0
Top decile	0	0	0

Table 38 Households type by decile as a proportion of all households of this type in York (2001, after policy changes)

Decile	Proportion of household type (%) Single person	Couple with no children	Couple with two children
Bottom decile	5.6	4.4	7.0
Decile 2	36.1	4.1	13.8
Decile 3	17.7	3.1	6.8
Decile 4	3.0	10.5	8.2
Decile 5	3.7	5.9	15.6
Decile 6	2.8	14.8	11.6
Decile 7	9.4	9.8	17.7
Decile 8	12.3	9.0	13.3
Decile 9	2.0	12.3	5.9
Top decile	7.3	26.2	0.0

Estimating the impact of welfare reforms introduced in April 2003

So far, the estimates of some of the major welfare reforms that were implemented in the late 1990s have been presented. Nevertheless, it should be noted that the Working Families' Tax Credit scheme, which was discussed above, was replaced in April 2003 by a new set of tax credits: the *Child Tax Credit (CTC)* and the *Working Tax Credit (WTC)*. The former aims at providing support for families in a common framework, in which the same rules apply to all households, whether in or out of work.[8] In particular, CTC can be claimed by all persons who are responsible for at least one child under 16 years of age or under 19 years and in full-time non-advanced education. CTC comprises five elements, which are listed in Table 39.

The CTC is calculated in a similar way to the WFTC. In particular, the family income is compared with the threshold figure, which is currently £13,230 per year, for those who do not claim WTC as well.[9] If the family income exceeds the threshold amount, the maximum CTC is reduced by 37 per cent of the excess.

Further, the Working Tax Credit (WTC) aims at providing a top-up to the wages of low-income workers. In particular, WTC can be claimed by all those with dependent children and/or a disability who work for 16 hours a week. Further, it can also be claimed by all those who do not have dependent children and do not have a disability, provided that they are aged 25 years or more and work at least 30 hours a week. The WTC elements are outlined in Table 40.

The WTC is calculated by comparing the maximum amount with the threshold figure,

Table 39 Child Tax Credits, weekly (April 2003)

Elements of Child Tax Credit	Amount in April 2003 (£)	Adjusted to 1991 prices (£)
Family element	10.45	7.38
Family element baby addition	10.45	7.38
Child element	27.75	19.61
Disabled child element	41.30	29.18
Enhanced disabled child element	16.60	11.73

Table 40 Working Tax Credits per week (April 2003)

Working Tax Credit	Amount in April 2003 (£)	Adjusted for 1991 prices (£)
Basic element	29.20	20.63
Couple or lone parent element	28.80	20.35
30 hours' credit	11.90	8.41
Disability element	39.15	27.66
Severe disability element	16.60	11.73
50-plus element	16.25	11.48
Childcare credit		
One child	135.00	95.39
Two or more children	200.00	141.31

which is £5,060 per year. As was the case with the CTC, if the income exceeds the threshold amount, the maximum WTC is reduced by 37 per cent of the excess. If a family claims both the WTC and the CTC, then the threshold amount to be compared with the maximum amount for all credits is £5,060 per year.

In the context of this book, the threshold amounts for the above credits were readjusted to their equivalent in 1991 on the basis of the RPI growth. Further, all the relevant elements of these credits were readjusted, before allocating them to eligible households of our simulated database. The actual (2003) and adjusted (1991) amounts for the various credits are shown in Tables 39 and 40.

Once all the figures were adjusted, the next step was to estimate the redistributive effects that the recently introduced policy reforms would have if they had been implemented in each of the simulation years, assuming full take-up. Table 41 summarises the estimated increase that would occur to the average incomes of households by class.

As can be seen by comparing Table 41 with Table 30, the new tax credits would result in a more significant increase of the average income of the poor and very poor households. For instance, in 1991, the increase in the income of the very poor households is estimated to more than double with the implementation of the new tax credits, compared to the respective increase presented in Table 30. Similar differences can be observed in all of the simulation years. These large differences may be explained by the fact that the Child Tax Credits can be claimed by unemployed individuals with children. Further, it should be noted that the Working Tax Credit can be claimed by individuals in poor households without children, whereas the previous credits under WFTC were aimed only at households with dependent children.

Estimating the geographical impacts

One of the major advantages of spatial microsimulation methods is the ability to estimate the impact of social and economic policies on different places. Further, it can be argued that identifying the geographical impact of national social policies is particularly important when these policies are aimed at breaking vicious circles of poverty in particular localities. As McCormick and Philo (1995) point out, much of the poverty in the UK is hidden, in the sense that *poor* people in average localities are largely invisible. Further, they argue that poverty in these localities is not only the result of economic decline reflected as a shift in demand for specific labour market skills, but also the cause of the decline. Moreover, Darton *et al*. (2003) point out that the distribution of disadvantage in the UK has a strongly geographical dimension.

> *A world of difference separates the poorest neighbourhoods where jobs are scarce and where poverty and associated social problems are the norm, from booming areas where the greatest difficulty is the shortage of staff for domestic and public services.*
> (Darton et al., 2003, p. 28)

However, as noted in Chapters 6 and 7, one of the difficulties associated with impact analysis at the small area level has been the lack of data pertaining to quality of life such as household income and wealth. Despite these difficulties, there have been a few attempts to address the geographical complexity of poverty

Table 41 Simulated impact of April 2003 policy changes by household class and simulation year

	Extra income (£) (in 1991 terms)	Extra income (£) (in 2003 terms)[a]	Income increase (%)	Income increase as % of all income in York
1991				
Very poor	10,365,947.06	13,403,169.55	34.1	1.90
Poor	5,199,802.76	6,723,344.97	10.02	0.95
Below average	4,911,779.79	6,350,931.27	7.82	0.90
Above average	2,454,956.07	3,174,258.19	3.59	0.45
Affluent	5,627,831.21	7,276,785.75	0.71	1.03
2001				
Very poor	10,689,087.19	13,820,989.74	29.3	1.64
Poor	6,511,030.14	8,418,761.97	9.07	1.00
Below average	4,169,833.85	5,391,595.17	6.62	0.00
Above average	4,570,505.07	5,909,663.06	5.00	0.00
Affluent	1,626,691.08	2,103,311.57	0.44	0.00
2011				
Very poor	11,173,514.24	14,447,353.91	26.0	1.46
Poor	8,108,874.62	10,484,774.88	11.05	1.06
Below average	7,747,491.80	10,017,506.91	10.52	1.01
Above average	3,945,888.96	5,102,034.43	3.27	0.52
Affluent	1,326,213.64	1,714,794.23	0.29	0.17
2021				
Very poor	16,409,094.65	21,216,959.38	27.15	1.99
Poor	5,904,514.42	7,634,537.15	11.35	0.72
Below average	11,121,892.69	14,380,607.25	4.90	1.35
Above average	4,661,651.58	6,027,515.49	2.65	0.57
Affluent	787,554.54	1,018,308.02	0.19	0.10

a Assuming that the growth of income for all household groups was equivalent to the RPI growth for the period 1991–2003.

(see, for instance, Dorling and Tomaney, 1995; Dorling and Woodward, 1996). Further, it has long been argued that spatial microsimulation can and should be used to fill in these data gaps in order to enable the analysis of the geographical dimensions of poverty and the evaluation of the spatial impacts of relevant social policies (Ballas and Clarke, 2001a).

The analysis presented so far is geographical in the sense that it describes the quality of life of households at the metropolitan district level (York). Clearly, this analysis can be extended to include all districts in Britain and to map socio-economic patterns across British regions and districts.

Nevertheless, it is also possible to use spatial microsimulation models to examine the impact of policy changes at the intra-district level. This section presents the geographical distribution of the simulated policy impacts within York. Figure 37 depicts the spatial distribution of the average additional household income that would result from the policy reforms discussed above. Moreover, Figure 38 depicts the spatial distribution of this additional income as a proportion of the average household income in each ward.

As can be seen, the areas of Bishophill, Guildhall and Fishergate would experience the highest average income increases, whereas the wards of Acomb, Foxwood and Micklegate would be the least affected localities in York. Further, as can be seen in Figure 38, the impact of the policy changes as a proportion of the average household income per ward will be greater on the wards of Westfield, Bishophill and Guildhall.

It is also interesting to examine what would have been the geographical impact of the new tax credits that were introduced in April 2003 and were briefly discussed above. Figure 39 shows the spatial distribution of the average additional household income that would result from the policy reforms discussed above, assuming that the April 2003 tax credits were implemented in 1991. Furthermore, Figure 40 depicts the spatial distribution of this additional income as a proportion of the average household income in each ward. As can be seen by comparing Figures 39 and 40 with Figures 37 and 38 respectively, the April 2003 changes in tax credits would result in relatively more income pumped into the areas of Bootham, Acomb and Walmgate. It is interesting to note that these areas had relatively high unemployment rates in 1991 and therefore they would be more likely to benefit from the April 2003 Child Tax Credits, which can be claimed by unemployed individuals with dependent children. It is also worth noting that these areas also had relatively high proportions of households with two or more dependent children in 1991.

The above geographical analysis can be repeated for different areas and at different geographical scales. Figures 41–4 illustrate the estimated geographical impact of the same policies in Wales in 1991, at the local authority district level.

Strengths and limitations

This chapter has demonstrated the potential of spatial microsimulation models for policy analysis. In particular, an estimate of the extent to which the simulated households would be affected by some of the policy changes that occurred in the late 1990s is made. The estimated impact of the recently introduced *Working Tax Credits* and *Child Tax Credits* is also presented. It should be noted that the analysis presented here is based on estimates and is aimed at providing an indication of the possible impacts of various policies. Further, this analysis did not take into account the behavioural responses of the population to the policy changes. For instance, there is no attempt to estimate how many unemployed family heads, who are able to work, would decide to enter the labour market as a result of welfare-to-work policies such as the Working Families' Tax Credit.

Figure 37 Estimated spatial distribution of additional income per household in 1991

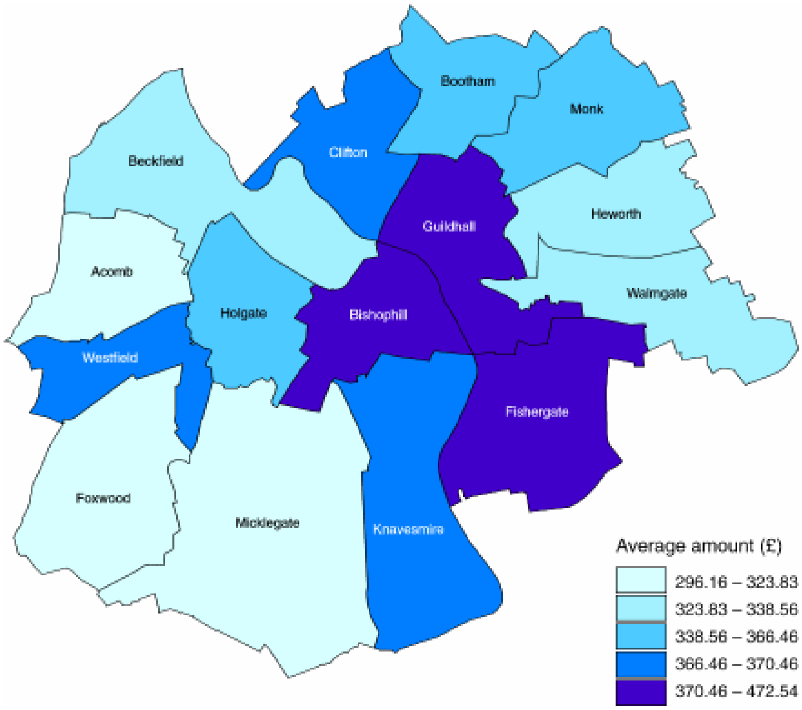


Figure 38 Spatial distribution of additional income per household as a proportion of average household income by ward in 1991

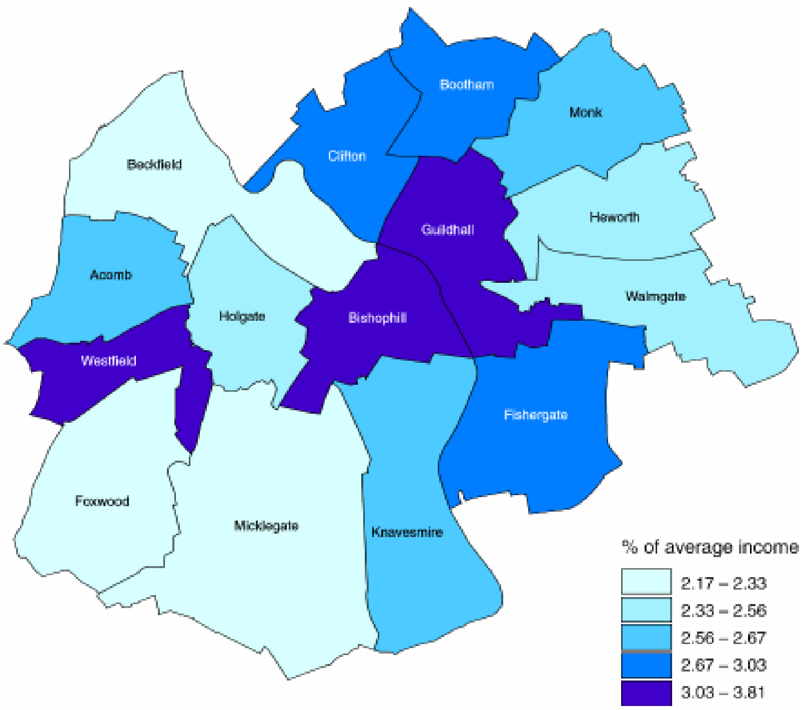


Figure 39 Estimated spatial distribution of additional income per household in 1991, after the implementation of the April 2003 Tax Credits

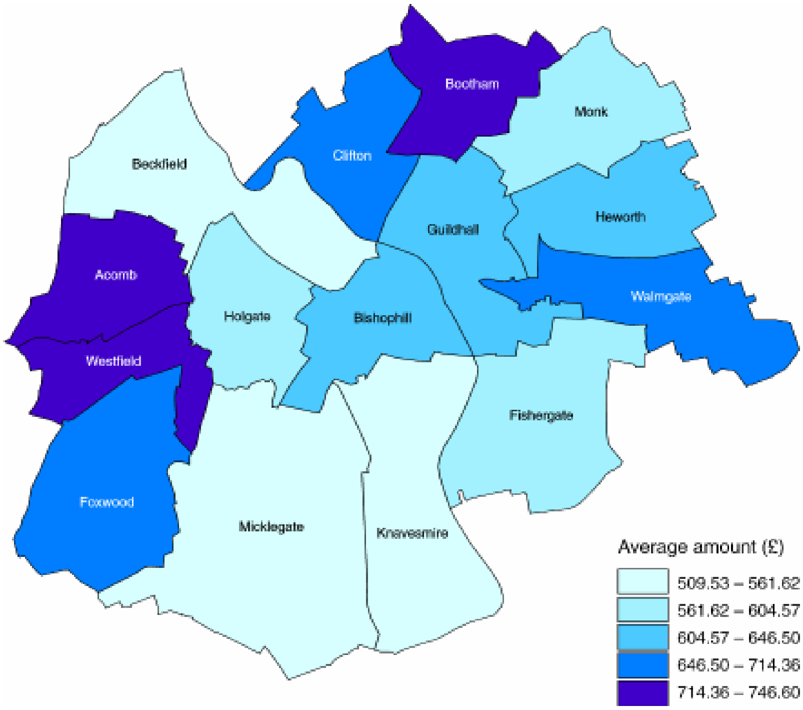


Figure 40 Spatial distribution of additional income per household as a proportion of average household income by ward, after the implementation of the April 2003 Tax Credits

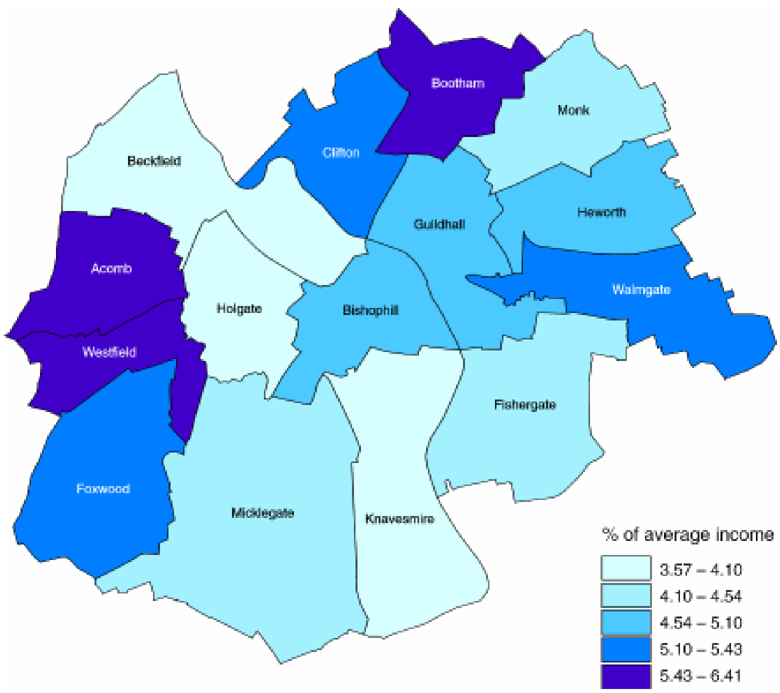


Figure 41 Estimated spatial distribution of additional income per household in 1991

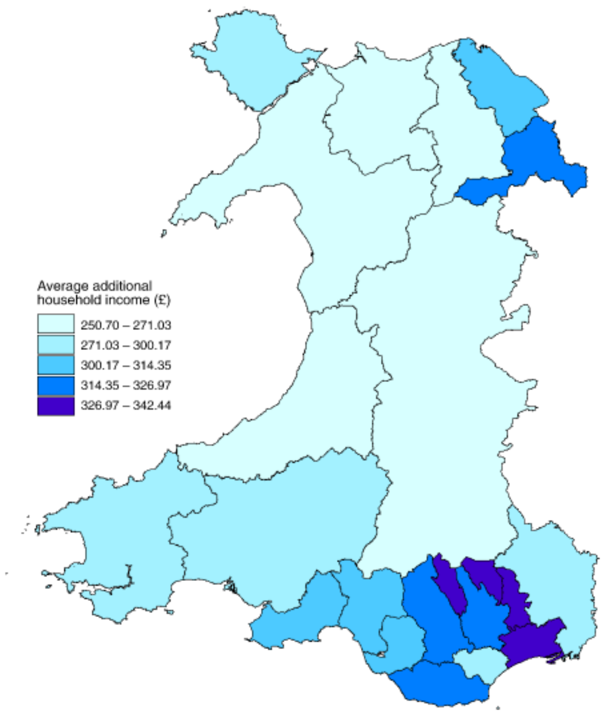


Figure 42 Spatial distribution of additional income per household as a proportion of average household income by district in 1991 (after implementing all policy changes described above)

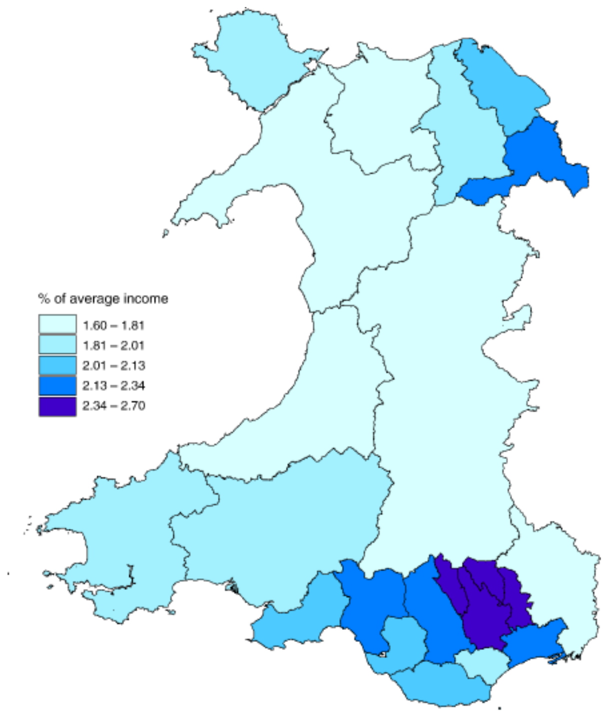


Figure 43 Estimated spatial distribution of additional income per household in 1991, after the implementation of the April 2003 tax credits

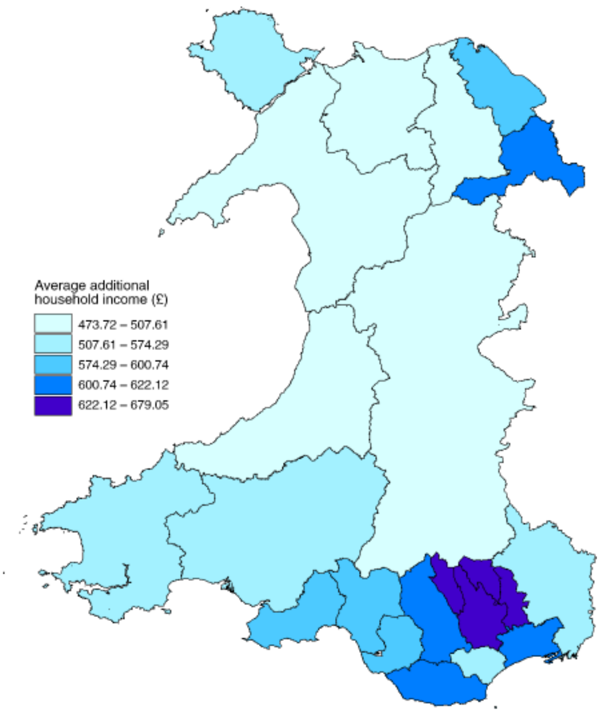
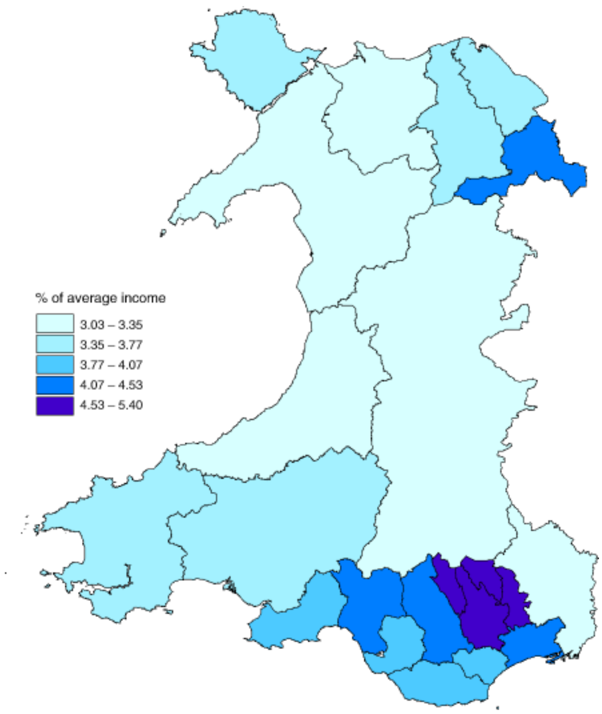


Figure 44 Spatial distribution of additional income per household as a proportion of average household income by ward, after the implementation of the April 2003 tax credits



15 Summary and conclusions

This book has presented a prototype method for the geographical analysis of national social policies. The method was developed on the basis of previous work in the area of microsimulation. It has been argued that the microsimulation models developed by economists have been very successful and widely used for policy analysis. Nevertheless, these models have not taken space into account for various reasons that include the paucity of geographical information on income, wealth, taxation and benefits and other suitable policy-relevant variables, as well as the perception that space is not important enough. This book has argued that space matters and that it is important to estimate the *geographical* as well as the *social*, *temporal* and *economic* impacts of policies. It has also been argued that spatial microsimulation methods can be used to estimate the geographical impacts of social policies. These can then be compared with the respective impacts of area-based policies, as *social policies* can be seen as alternatives to *area-based policies*. Further, it can be argued that spatial microsimulation methods can be used to analyse social policy in a geographically oriented *proactive* fashion. For instance, spatial microsimulation can be employed to identify deprived localities in which poor individuals and households are over-represented. Spatial microsimulation modelling can then be used to answer questions such as: 'What social policy could be applied, which, all else being equal would most likely improve the quality of life of residents in the inner-city localities of a city?'. In other words, new social policies can be formulated on the basis of spatial microsimulation outputs. Spatially oriented social policies can be seen as a substitute or an alternative to traditional area-based policies and direct comparisons of their efficiency and effectiveness can be made. Another interesting research area in which spatial microsimulation models can play a significant role is the measurement and analysis of income and wealth distribution. It would be useful to create a new set of deprivation indicators based on household poverty, wealth and social exclusion that can replace or be used in tandem with traditional census-based proxies for income and wealth such as car ownership and household tenure. One of the ways of evaluating the impact of social policy on the welfare of society is to compute indices such as income inequality measures and then to study the change of these indices after a social policy programme is applied. It can be argued that the mapping of inequality measures can highlight the degree of heterogeneity of income, household types and lifestyles within small areas.

This book has proposed a method of estimating policy-relevant geographical information at various spatial scales. Further, it has argued for a geographical approach to national policy analysis by presenting a prototype spatial microsimulation method that can be used to estimate the geographical distribution of a wide range of extremely policy-relevant social and economic variables. It is hoped that the method was presented in a simple enough way (with an attempt to keep statistical jargon to a minimum) to be understood and implemented by social scientists and policy practitioners with a basic understanding of quantitative methods. This book has demonstrated how it is possible to apply the geographical microsimulation method using a combination of publicly available socio-economic datasets. In addition, it has shown how the outputs of the simulation method can be used for policy analysis. Illustrative examples were presented and the strengths and limitations of the approach were identified.

It should be noted that the spatial microsimulation method proposed and implemented in this book has a great deal of potential for further improvement. The method as it stands at the moment can be used to paint a picture of one possible future of a city or region, based on past trends. In addition, the method is suitable for the estimation of variables such as household income at the small area level. Such estimations can provide helpful insights into the analysis of spatial and socio-economic polarisation within cities and regions. The simulation method can also be used to paint a picture of the life of households of different income categories. In this respect, the simulation outputs are very similar to large-scale survey outputs and qualitative research findings. The method is also useful in modelling the socio-economic and spatial effects of policy change. In particular, in the context of this book, the effects of policy changes over the last ten years were investigated (including the 2003 new Tax and Working Credits and the Family Credit, which they replaced). However, these estimates can only be used as an indication of the initial impact of policy change on different types of households in various localities. The method as it stands at the moment cannot be used to analyse the longer-term behavioural responses to policy changes. For instance, it cannot be used, in its current form, to predict how many unemployed individuals would decide, if they could, to enter the labour market as a result of increases in the Minimum Wage, or welfare-to-work policies such as the Working Families' Tax Credit.

It should be noted that, as described in Chapter 12, the method performs better at the district and parliamentary constituency level, than at the ward level. It is therefore more suitable for the prediction of a wide range of socio-economic variables at the coarser geographical level of cities and regions, but is less suitable for analysis of most variables at small area levels such as wards and enumeration districts. At the very local level, geography matters too much to be simulated easily!

It should also be noted that the geographical simulation method is not suitable for the prediction of rare or badly reported events, such as drug use. Also, it is unsuitable for the prediction of variables that are affected considerably by external and localised factors, such as transport networks and public transport services, or the presence of a proportionally large university or a single major employer in the region. However, a method of projecting variables into the future that does take confounding into consideration was presented in Chapter 13.

Overall, it can be argued that the geographical microsimulation method presented in this book can be used to provide useful information on socio-economic trends, as well as on the possible outcome of policy reforms, at different (albeit coarse) geographical scales. It can be argued that the analyses presented in this book could be used to stimulate debate about the future.

It is hoped that this book will highlight the importance of taking geography into account in national policy making. It is also hoped that the method and related conceptual and technical issues that were discussed in the book will lead to a more systematic discussion between economists and geographers, and that it will promote a convergence in the use of some methods between the very many different traditions that populate this field.

Notes

Chapter 1

1 Microsimulation was first introduced by Orcutt (1957).

Chapter 3

1 URL: http://www.smc.kiruna.se.

Chapter 9

1 In the context of this book, this process was applied iteratively on five tables and convergence was achieved after ten iterations.

Chapter 10

1 These equations are simple Newtonian projections in log space taking into account momentum and acceleration reprojected into real space.

Chapter 14

1 As Atkinson (1983) puts it, the 'two can live as cheaply as one' hypothesis.

2 It should be emphasised that we used households of the first BHPS wave (1991) in the simulation of all years.

3 DTI (online) *Employment Relations*: http://www.dti.gov.uk/er/nmw/.

4 http://www.inlandrevenue.gov.uk/.

5 http://www.bbc.co.uk/pressoffice/keyfacts/stories/licencefee.shtml.

6 For example, Family Credit was replaced by WFTC in 1999. The estimates presented here did not take the Family Credit amounts out of the original household incomes of eligible households. Therefore, the real effect of the policy changes will be less than that presented here. Nevertheless, it should be noted that, under the previous Family Credit scheme, there were far fewer eligible households compared to the WFTC scheme.

7 In order to carry out these calculations, the RPI growth rate was used to readjust the simulated household incomes for 2001.

8 Inland Revenue website: http://www.inlandrevenue.gov.uk/taxcredits/changes.htm#ctc.

9 http://www.inlandrevenue.gov.uk/rates/taxcredits.htm.

10 http://www.inlandrevenue.gov.uk/rates/taxcredits.htm.

References

Armstrong, H. and Taylor, J. (1993) *Regional Economics and Policy*. Brighton: Harvester Wheatsheaf

Atkinson, A.B. (1983) *The Economics of Inequality*. Oxford: Clarendon Press

Ballas, D. (2001) 'A spatial microsimulation approach to local labour market policy analysis', unpublished PhD thesis, School of Geography, University of Leeds

Ballas, D. and Clarke, G.P. (2000) 'GIS and microsimulation for local labour market policy analysis', *Computers, Environment and Urban Systems*, Vol. 24, pp. 305–30

Ballas, D. and Clarke, G.P. (2001a) 'Modelling the local impacts of national social policies: a spatial microsimulation approach', *Environment and Planning C: Government and Policy*, Vol. 19, pp. 587–606

Ballas, D. and Clarke, G.P. (2001b) 'Towards local implications of major job transformations in the city: a spatial microsimulation approach', *Geographical Analysis*, Vol. 33, pp. 291–311

Ballas, D., Clarke, G.P. and Commins, P. (2001) 'Building a spatial microsimulation model for the Irish rural economy', paper presented at the 12th Theoretical and Quantitative Geography European Colloquium, Rouen, France, 7–11 September

Ballas, D., Clarke, G.P. and Turton, I. (1999) 'Exploring microsimulation methodologies for the estimation of household attributes', paper presented at the 4th International Conference on GeoComputation, Fredericksburg, VA, 25–28 July

Ballas, D., Clarke, G.P., Dorling, D., Eyre, H., Rossiter, D. and Thomas, B. (2003) 'SimYork: simulating current and future trends in the life of households in York', report to the Joseph Rowntree Foundation, May

Barker, T. and Peterson, W. (eds) (1987) *The Cambridge Multisectoral Dynamic Model of the British Economy*. Cambridge: Cambridge University Press

Barker, T., Fingleton, B., Homenidou, K., Lewney, R. (2001) 'The regional Cambridge multisectoral dynamic model of the UK economy', in G.P. Clarke and M. Madden (eds) *Regional Science in Business*. Berlin: Springer-Verlag, pp. 79–96

Batey, P.W.J. and Madden, M. (1999) 'The employment impact of demographic change: a regional analysis', *Papers in Regional Science*, Vol. 78, pp. 69–87

Batey, P.W.J. and Madden, M. (2001) 'Socio-economic impact assessment: meeting client requirements', in G.P. Clarke and M. Madden (eds) *Regional Science in Business*. Berlin: Springer-Verlag, pp. 37–60

Bertuglia, C.S., Clarke, G.P. and Wilson, A.G. (1994) 'Models and performance indicators in urban planning: the changing policy context', in C.S. Bertuglia, G.P. Clarke and A.G. Wilson (eds) *Modelling the City*. London: Routledge, pp. 20–36

Birkin, M. and Clarke, M. (1988) 'SYNTHESIS – a synthetic spatial information system for urban and regional analysis: methods and examples', *Environment and Planning A*, Vol. 20, pp. 1645–71

Birkin, M. and Clarke, M. (1989) ‘The generation of individual and household incomes at the small area level using Synthesis’, *Regional Studies*, Vol. 23, pp. 535–48

Birkin, M. and Clarke, G.P. (1995) ‘Using microsimulation methods to synthesize census data’, in S. Openshaw (ed.) *Census Users’ Handbook*. London: GeoInformation International, pp. 363–87

Birkin, M., Clarke G.P. and Clarke, M. (1996) ‘Urban and regional modelling at the microscale’, in G.P. Clarke (ed.) *Microsimulation for Urban and Regional Policy Analysis*. London: Pion, pp. 10–27

Brimblecombe, N., Dorling, D. and Shaw, M. (2000) ‘Migration and geographical inequalities in health in Britain: an exploration of the lifetime socio-economic characteristics of migrants’, *Social Science and Medicine*, Vol. 50, pp. 861–78

Burrows, R. and Rhodes, D. (1998) *Unpopular Places?: Area Disadvantage and the Geography of Misery in England*. Bristol: Polity Press

Caldwell, S.B. and Keister, L.A. (1996) ‘Wealth in America: family stock ownership and accumulation, 1960–1995’, in G.P. Clarke (ed.) *Microsimulation for Urban and Regional Policy Analysis*. London: Pion, pp. 88–116

Caldwell, S.B., Clarke, G.P. and Keister, L.A. (1998) ‘Modelling regional changes in US household income and wealth: a research agenda’, *Environment and Planning C: Government and Policy*, Vol. 16, pp. 707–22

Clarke, G.P. (1996) ‘Microsimulation: an introduction’, in G.P. Clarke (ed.) *Microsimulation for Urban and Regional Policy Analysis*. London: Pion, pp. 1–9

Clarke, G.P. and Wilson, A.G. (1994) ‘A new geography of performance indicators for urban planning’, in C.S. Bertunglia, G.P. Clarke and A.G. Wilson (eds) *Modelling the City*. London: Routledge, pp. 55–81

Clarke, M. (1986) ‘Demographic processes and household dynamics: a microsimulation approach’, in R. Woods and P.H. Rees (eds) *Population Structures and Models: Developments in Spatial Demography*. London: Allan & Unwin, pp. 245–72

Cole, K. (1993) ‘The 1991 Local Base and Small Area Statistics’, in A. Dale and C. Marsh (eds) *The 1991 Census User’s Guide*. London: HMSO, pp. 201–47

Connolly, C. and Chisholm, M. (1999) ‘The use of indicators for targeting public expenditure: the Index of Local Deprivation’, *Environment and Planning C: Government and Policy*, Vol. 17, pp. 463–82

Coombes, M. (1995) ‘Dealing with census geography: principles, practices and possibilities’, in S. Openshaw (ed.) *Census Users’ Handbook*. London: GeoInformation International, pp. 111–32

Curtis, S. (1995) ‘Geographical perspectives on poverty, health and health policy in different parts of the UK’, in C. Philo (ed.) *Off the Map: The Social Geography of Poverty in the UK*. London: Child Poverty Action Group, pp. 153–88

Dale, A. (1993) 'The content of the 1991 Census: change and continuity', in A. Dale, C. Marsh (eds) *The 1991 Census User's Guide*. London: HMSO, pp. 16–51

Dale, A. and Teague, A. (2002) 'Microdata from the Census: samples of anonymised records', in P. Rees, D. Martin and P. Williamson (eds) *The Census Data System*. Chichester: Wiley, pp. 203–12

Dale, A., Fieldhouse, E. and Holdsworth, C. (2000) *Analyzing Census Microdata*. London: Arnold

Darton, D., Hirsch, D. and Strelitz, J. (2003) *Tackling Disadvantage: A 20-year Enterprise*. YPS for the Joseph Rowntree Foundation

Denham, C. (1993) 'Census geography: an overview', in A. Dale and C. Marsh (eds) *The 1991 Census User's Guide*. London: HMSO, pp. 52–69

Dhanecha, N., Ellerd-Elliott, S., Herring, I., Horsfall, E., Majetic, P., Shome, J. and Snow, J. (2003) *Family Resources Survey: Great Britain 2001–02*. London: Department for Work and Pensions, available online at http://www.dwp.gov.uk/asd/frs/2001_02/pdfonly/frs_2001_02_report.pdf

Dixie, J. and Dorling, D. (2002) 'New questions for the 2001 Census', in P. Rees, D. Martin and P. Williamson (eds) *The Census Data System*. Chichester: Wiley, pp. 283–94

Dorling, D. (1994) 'The negative equity map of Britain', *Area*, Vol. 26, No. 4, pp. 327–42

Dorling, D. and Tomaney, J. (1995) 'Poverty in the old industrial regions: a comparative view', in C. Philo (ed.) *Off the Map: The Social Geography of Poverty in the UK*. London: Child Poverty Action Group, pp. 103–32

Dorling, D. and Woodward, R. (1996) 'Social polarisation 1971–1991: a micro-geographical analysis of Britain', *Progress in Planning*, Vol. 45, No. 2, pp. 1–67

Duley, C.J. (1989) 'A model for updating Census-based household and individual information for inter-censal years', unpublished PhD thesis, School of Geography, University of Leeds

Duley, C.J., Rees, P.H. and Clarke, M.C. (1988) *A Microsimulation Model for Updating Households in Small Areas between Censuses*. Working Paper 515. Leeds: School of Geography, University of Leeds

Edwards, J. (1995) 'Social policy and the city: a review of recent policy developments and literature', *Urban Studies*, Vol. 32, pp. 695–712

Ernst, L.R., Hubble, D.L. and Judkins, D.R. (1984) 'Longitudinal family and household estimation in SIPP', *Journal of the American Statistical Association*, available online at http://www.amstat.org/sections/srms/Proceedings/papers/1984_129.pdf)

Essex Data Archive (2000) *Household Expenditure Survey*, available online at http://www.data-archive.ac.uk/findingData/fesAbstract.asp

Essex Data Archive (2003) *ONS Omnibus Survey*, available online at http://www.data-archive.ac.uk/findingData/omnibusTitles.asp

Evandrou, M. and Falkingham, J. (1995) ‘Gender, lone-parenthood and lifetime incomes’, in J. Falkingham and J. Hills (eds) *The Dynamic of Welfare: The Welfare State and the Life Cycle*. New York: Prentice Hall/Harvester Wheatsheaf, pp. 167–83

Falkingham, J. and Hills, J. (1995a) ‘The effects of the welfare state over the life cycle’, in J. Falkingham and J. Hills (eds) *The Dynamic of Welfare: The Welfare State and the Life Cycle*. New York: Prentice Hall/Harvester Wheatsheaf, pp. 83–107

Falkingham, J. and Hills, J. (1995b) ‘Redistribution between people or across the life cycle?’, in J. Falkingham and J. Hills (eds) *The Dynamic of Welfare: The Welfare State and the Life Cycle*. New York: Prentice Hall/Harvester Wheatsheaf, pp. 137–49

Falkingham, J. and Lessof, C. (1992) ‘Playing God or LIFEMOD – the construction of a dynamic microsimulation model’, in R. Hancock and H. Sutherland (eds) *Microsimulation Models for Public Policy Analysis: New Frontiers*. London: Suntory-Toyota International Centre for Economics and Related Disciplines, LSE, pp. 5–32

Falkingham, J., Harding, A. and Lessof, C. (1995) ‘Simulating lifetime income distribution and redistribution’, in J. Falkingham and J. Hills (eds) *The Dynamic of Welfare: The Welfare State and the Life Cycle*. New York: Prentice Hall/Harvester Wheatsheaf, pp. 62–82

Fitzpatrick, P. *et al*. (2002) *Welfare Benefits Handbook 2002/03*. London: Child Poverty Action Group

Frick, J.R. and Haisken-DeNew, J.P. (2001) *Structuring the HILDA Panel: Considerations and Suggestions*. Hilda Project discussion paper series. Melbourne: Dept of Family and Community Services, University of Melbourne, available online at http://www.melbourneinstitute.com/hilda/hdpsn02.pdf)

Gilbert, N. and Troitzsch, K.G. (1999) *Simulation for the Social Scientist*. Buckingham: Open University Press

Glennerster, H., Falkingham, J. and Barr, N. (1995) ‘Education funding, equity and the life cycle’, in J. Falkingham and J. Hills (eds) *The Dynamic of Welfare: The Welfare State and the Life Cycle*. New York: Prentice Hall/Harvester Wheatsheaf, pp. 150–66

Goodwin, M. (1995) ‘Poverty in the city: “you can raise your voice, but who is listening?”’, in C. Philo (ed.) *Off the Map: The Social Geography of Poverty in the UK*. London: Child Poverty Action Group, pp. 65–82

Gordon, D. and Pantazis, C. (1997) ‘Measuring poverty: breadline Britain in the 1990s’, in D. Gordon and C. Pantazis (eds) *Breadline Britain in the 1990s*. Aldershot: Ashgate, pp. 1–47

Green, A. (1996) ‘Aspects of the changing geography of poverty and wealth’, in J. Hills (ed.) *New Inequalities: The Changing Distribution of Income and Wealth in the United Kingdom*. Cambridge: Cambridge University Press, pp. 265–91

Hancock, R. (2000) 'Charging for care in later life: an exercise in dynamic microsimulation', in L. Mitton, H. Sutherland, and M. Weeks (eds) *Microsimulation Modelling for Policy Analysis: Challenges and Innovations*. Cambridge: Cambridge University Press, pp. 226–37

Hancock, R. and Sutherland, H. (eds) (1992) *Microsimulation Models for Public Policy Analysis: New Frontiers*. London: Suntory-Toyota International Centre for Economics and Related Disciplines, LSE

Hancock, R., Mallender, J. and Pudney, S. (1992) 'Constructing a computer model for simulating the future distribution of pensioners' incomes for Great Britain', in R. Hancock and H. Sutherland (eds) *Microsimulation Models for Public Policy Analysis: New Frontiers*. London: Suntory-Toyota International Centre for Economics and Related Disciplines, LSE, pp. 33–66

Harding, A. (ed.) (1996) *Microsimulation and Public Policy*. Contributions to Economic Analysis No. 232. Amsterdam: North Holland

Holm, E., Lindgren, U., Makila, K. and Malmberg, G. (1996) 'Simulating an entire nation', in G.P. Clarke (ed.) *Microsimulation for Urban and Regional Policy Analysis*. London: Pion, pp. 164–86

Hooimeijer, P. (1996) 'A life-course approach to urban dynamics: state of the art in and research design for the Netherlands', in G.P. Clarke (ed.) *Microsimulation for Urban and Regional Analysis*. London: Pion, pp. 28–63

Isard, W. (1960) *Methods of Regional Analysis: An Introduction to Regional Science*. Cambridge, MA: MIT Press/Wiley

Isard, W. (1975) *Introduction to Regional Science*. London: Prentice Hall

Jin, Yu-xian and Wilson, A.G. (1993) 'Generation of integrated multispatial input-output models of cities (GIMIMoC): 1: initial stage', *Papers in Regional Science*, Vol. 72, pp. 351–68

Johnston, R., Pattie, C., Rossiter, D., Dorling, D., Tunstall, H. and MacAllister, I. (1998) 'Anatomy of a Labour landslide: the constituency system and the 1997 general election', *Parliamentary Affairs*, Vol. 51, No. 2, pp. 131–48

Leontief, W. (1951) *The Structure of the American Economy: 1919–1929*. New York: Oxford University Press

Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (eds) (1999) *Geographical Information Systems: Management Issues and Applications*. New York: Wiley

Madden, M. (1993) 'Welfare payments and migration in a nonlinear, extended input-output model with an application to Scotland', *Papers in Regional Science*, Vol. 72, pp. 177–99

Marsh, C. (1993) 'The sample of anonymised records', in A. Dale and C. Marsh (eds) *The 1991 Census User's Guide*. London, HMSO, pp. 295–311

Marsh, C. and Teague, A. (1992) 'Samples of anonymised records from the 1991 Census', *Population Trends*, Vol. 69, pp. 17–26

Martin, D. (1996) *Geographic Information Systems: Socioeconomic Applications*. London: Routledge

Martin, R. (1995) 'Income and poverty inequalities across regional Britain: the North–South divide lingers on', in C. Philo (ed.) *Off the Map: The Social Geography of Poverty in the UK*. London: Child Poverty Action Group, pp. 23–44

McCormick, J. and Philo, C. (1995) 'Where is poverty? The hidden geography of poverty in the United Kingdom', in C. Philo (ed.) *Off the Map: The Social Geography of Poverty in the UK*. London: Child Poverty Action Group, pp. 1–22

McKendrick, J. (1995) 'Poverty in the UK: the Celtic divide', in C. Philo (ed.) *Off the Map: The Social Geography of Poverty in the UK*. London: Child Poverty Action Group, pp. 45–64

Mertz, J. (1991) 'Microsimulation – a survey of principles developments and applications', *International Journal of Forecasting*, Vol. 7, pp. 77–104

Middleton, E. (1995) 'Samples of anonymized records', in S. Openshaw (ed.) *Census Users' Handbook*. London: GeoInformation International, pp. 337–62

MIMAS (Manchester Information and Associated Services) (2000) *MIMAS Datasets and Availability*, available online at http://www.mimas.ac.uk/datasets.html

Mitchell, R., Dorling, D. and Shaw, M. (2000) *Inequalities in Life and Death: What if Britain were more Equal?* Bristol: The Policy Press

Mitton, L, Sutherland, H. and Weeks, M. (eds) (2000) *Microsimulation Modelling for Policy Analysis: Challenges and Innovations*. Cambridge: Cambridge University Press

Noble, M. and Smith, G. (1996) 'Two nations? Changing patterns of income and wealth in two contrasting areas', in J. Hills (ed.) *New Inequalities: The Changing Distribution of Income and Wealth in the United Kingdom*. Cambridge: Cambridge University Press, pp. 292–318

Noble, M., Smith, G., Wright, G., Dibben, C. and Lloyd, M. (2001) *Measures of Deprivation in Northern Ireland*. Belfast: Northern Ireland Statistical Research Agency (NISRA), available online at http://www.nisra.gov.uk/whatsnew/dep/index.html

ONS (Office for National Statistics) (1997) *New Earnings Survey 1997*. London: The Stationery Office

ONS (Office for National Statistics) (2000a) *Family Spending: A Report on the 1998–99 Family Expenditure Survey*. London: The Stationery Office

ONS (Office for National Statistics) (2000b) *Family Resources Survey: Great Britain*. London: Department of Social Security

ONS (Office for National Statistics) (2001a) *New Earnings Survey – United Kingdom: Streamlined Analyses; Description of the Survey*, available online at http://www.statistics.gov.uk/statbase/Product.asp?vlnk=647

ONS (Office for National Statistics) (2001b) *Labour Force Survey*, available online at http://www.statistics.gov.uk/themes/labour_market/surveys/labour_force_text.asp

ONS (Office for National Statistics) (2001c) *Family Expenditure Survey*, available online at http://www.statistics.gov.uk/themes/economy/Surveys/survey_of_fes.asp

ONS (Office for National Statistics) (2003a) *General Household Survey*, available online at http://www.statistics.gov.uk/ssd/surveys/general_household_survey.asp

ONS (Office for National Statistics) (2003b) *National Statistics Omnibus Survey*, available online at http://www.statistics.gov.uk/STATBASE/Source.asp?vlnk=657&More=Y

ONS (Office for National Statistics) (2003c) *Labour Force Survey*, available online at http://www.statistics.gov.uk/ssd/surveys/labour_force_survey.asp

ONS (Office for National Statistics) (2003d) *Guide to RPI and Finding Data FAQs on the UK's Most Familiar Measure of Inflation*, available online at http://www.statistics.gov.uk/cci/nugget.asp?id=21

Openshaw, S. (ed.) (1995) *Census Users' Handbook*. London: GeoInformation International

Orcutt, G.H. (1957) 'A new type of socio-economic system', *The Review of Economics and Statistics*, Vol. 39, pp. 116–23

Orcutt, G.H., Mertz, J. and Quinke, H. (eds) (1986) *Microanalytic Simulation Models to Support Social and Financial Policy*. Amsterdam: North-Holland

Propper, C. (1995) 'For richer, for poorer, in sickness and in health: the lifetime distribution of NHS health care', in J. Falkingham and J. Hills (eds) *The Dynamic of Welfare: The Welfare State and the Life Cycle*. New York: Prentice Hall/Harvester Wheatsheaf, pp. 184–203

Redmond, G., Sutherland, H. and Wilson, M. (1998) *The Arithmetic of Tax and Social Security Reform: A User's Guide to Microsimulation Methods and Analysis*. Cambridge: Cambridge University Press

Rees, P. (1995) 'Putting the census on the researcher's desk', in S. Openshaw (ed.) *Census Users' Handbook*. London: GeoInformation International, pp. 27–82

Rees, P., Martin, D. and Williamson, P. (2002) 'Census data resources in the United Kingdom', in P. Rees, D. Martin and P. Williamson (eds) *The Census Data System*. Chichester: Wiley, pp. 1–24

Rowntree, B.S. (2000) *Poverty: A Study of Town Life*. Bristol: Policy Press

Rogerson, P.A. and Plane, D.A. (1998) 'The dynamics of neighborhood age composition', *Environment and Planning A*, Vol. 30, pp. 1461–72

Scholten, H. and Stillwell, J.C.H. (eds) (1990) *Geographic Information Systems for Urban and Regional Planning*. Dordrecht: Kluwer

Shephard, A. (2003) *Inequality under the Labour Government*. Briefing Note 33. London: The Institute for Fiscal Studies

Stillwell, J. and Scholten, H. (eds) (2001) *Land Use Simulation for Europe*. Dordrecht: Kluwer

Sutherland, H. and Piachaud, D. (2001) 'Reducing child poverty in Britain: an assessment of government policy 1997–2001', *The Economic Journal*, Vol. 111, pp. 85–101

Sutherland, H., Sefton, T. and Piachaud, D. (2003) *Poverty in Britain: The Impact of Government Policy since 1997*. York: Joseph Rowntree Foundation, York, also available online at http://www.jrf.org.uk

Swan, N. (2000) 'Unemployment insurance and labour mobility: analysis using a new Swedish microsimulation model', in L. Mitton, H. Sutherland and M. Weeks (eds) *Microsimulation Modelling for Policy Analysis: Challenges and Innovations*. Cambridge: Cambridge University Press, pp. 251–67

Taylor, M.F. (ed). with Brice, J., Buck, N. and Prentice-Lane, E. (2001) *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. Colchester: University of Essex

Vencatasawmy, C.P., Holm, E. and Rephann, T. *et al*. (1999) 'Building a spatial microsimulation model', paper presented at the 11th Theoretical and Quantitative Geography European Colloquium, Durham Castle, Durham, 3–7 September

Williamson, P. (1992) 'Community care policies for the elderly: a microsimulation approach', unpublished PhD thesis, School of Geography, University of Leeds

Williamson, P., Birkin, M. and Rees, P. (1998) 'The estimation of population microdata by using data from small area statistics and samples of anonymised records', *Environment and Planning A*, Vol. 30, pp. 785–816

Appendix

A regression framework for continuous and binary variables

This appendix builds on the discussion presented in Chapter 12 and shows how it is possible to build a regression framework for both continuous variables (such as household income) and binary variables (such as PC ownership).

Continuous variables

Initial modelling involved the specification of ten equations for each dependent variable, one for each wave. This allowed not only the general level of the dependent variable to change over time, but also the pattern of dependence. The next task was to use the information from the ten models to project patterns into the future. In the course of trying to devise an appropriate methodology, it became clear that an alternative, conceptually far simpler, approach was also available.

Consider the following set of simplified equations for the first three waves of the BHPS:

LOG(INCOME wave a) = a + b11REGION + b12CAR + b13TENURE + …..

LOG(INCOME wave b) = a + b21REGION + b22CAR + b23TENURE + …..

LOG(INCOME wave c) = a + b31REGION + b32CAR + b33TENURE + …..

If an additional variable, WAVE, is specified, then data from waves a, b and c can be combined within the same model as follows:

LOG(INCOME) = a + b0WAVE + b1REGION + b2CAR + b3TENURE + …..

The advantage of this model is that it pools the data from the three waves, thereby increasing the reliability of the parameter estimates. The disadvantage is that the effect of each dependent variable is held constant over time. Thus, whereas the first approach allows for changing patterns of relationship between dependent and independent variables (for example, the increasing gap in rates of PC ownership between car owners and non-car owners over the course of the ten waves), the second approach does not. The solution to this is to allow interaction effects between WAVE and each of the constraint variables in the model:

LOG(INCOME) = a + b0WAVE + b1REGION + b2CAR + b3TENURE + ….. + b11WAVE × REGION + b12WAVE × CAR + b13WAVE × TENURE + …..

This gives two advantages: the increased reliability of parameter estimates resulting from the pooling of data together with a model that allows changing patterns of dependence on the independent variables over time. The parameter estimates for modelling both log of household income and logit of PC ownership are shown in Table A.1. In terms of the above equation, the coefficients a, b1, b2, b3 … appear in the column main effect; while the coefficients b0, b11, b12, b13 … appear in the column wave effect.

Table A.1 Parameter estimates for log (household income) and logit (PC ownership) together with example estimates for the period 1991–2020

Income intercept/ wave	Main effect 9.10	Wave effect 0.014	Example estimates		PC ownership intercept/wave	Main effect -1.51	Wave effect 0.125	Example estimates	
			1991	19015				1991	0.40
			1992	19264				1992	0.43
London	0.14	0.014	1993	19516	London	0.24	0.021	1993	0.45
SE	-0.04	0.018	1994	19772	SE	0.07	0.006	1994	0.47
SW	-0.19	0.022	1995	20030	SW	-0.10	0.013	1995	0.49
Wales	0.06	-0.016	1996	20292	Wales	-0.16	0.022	1996	0.52
E Anglia	-0.18	0.013	1997	20558	E Anglia	-0.24	0.014	1997	0.54
E Midlands	-0.10	0.009	1998	20827	E Midlands	0.16	-0.032	1998	0.56
W Midlands	-0.13	0.011	1999	21099	W Midlands	-0.23	0.010	1999	0.58
NW	-0.07	0.020	2000	21375	NW	0.10	-0.016	2000	0.61
Yorks & Humber	-0.11	0.012	2001	21655	Yorks & Humber	-0.11	-0.007	2001	0.63
North	-0.04	0.010	2002	21939	North	0.18	-0.032	2002	0.65
Scotland	0.00	0.000	2003	22226	Scotland	0.09	0.001	2003	0.67
No car	-0.76	-0.002	2004	22516	No car	-0.41	-0.037	2004	0.69
One car	-0.36	-0.001	2005	22811	One car	-0.04	0.006	2005	0.71
Two+ cars	0.00	0.000	2006	23110	Two+ cars	0.45	0.031	2006	0.73
Owned	0.25	0.022	2007	23412	Owned	0.20	-0.001	2007	0.74
Social rented	-0.03	0.030	2008	23718	Social rented	-0.28	-0.024	2008	0.76
Other rented	0.00	0.000	2009	24029	Other rented	0.08	0.025	2009	0.78
High status	0.63	0.007	2010	24343	High status	0.48	0.024	2010	0.79
Medium status	0.40	-0.001	2011	24662	Medium status	0.31	-0.015	2011	0.81
Low status	0.41	0.001	2012	24984	Low status	0.10	-0.013	2012	0.82
Retired	-0.10	0.010	2013	25311	Retired	-0.95	-0.008	2013	0.83
Inactive	0.00	0.000	2014	25642	Inactive	0.06	0.012	2014	0.85
Married	0.39	-0.005	2015	25978	Married	0.18	-0.011	2015	0.86
Single parent	0.13	-0.004	2016	26318	Single parent	0.16	-0.006	2016	0.87
Other	0.00	0.000	2017	26662	Other	-0.34	0.017	2017	0.88

(Continued overleaf)

Table A.1 Parameter estimates for log (household income) and logit (PC ownership) together with example estimates for the period 1991–2020 (Continued)

Income intercept/ wave	Main effect 9.10	Wave effect 0.014	Example estimates		PC ownership intercept/wave	Main effect -1.51	Wave effect 0.125	Example estimates	
No children	0.10	-0.016	2018	27011	No children	-0.71	0.036	2018	0.89
1 child	0.06	-0.005	2019	27364	1 child	0.17	-0.001	2019	0.90
2+ children	0.00	0.000	2020	27723	2+ children	0.54	-0.035	2020	0.90

As an example of how the estimated income of a household is calculated, consider a Welsh household in wave c (the third wave), with one car, living in an owner-occupied house, with a medium-status head of household, married with two children. The intercept, six main effects, wave effect and six interaction terms are added (each of the last two multiplied by three because this is the third wave, though as two+ children is the [arbitrary] reference category its effects are zero) and the resulting value exponentiated:

INCOME = EXP (9.10 + 0.06 - 0.36 + 0.25 + 0.40 + 0.39 + 0.00 + 0.042 - 0.048 - 0.003 + 0.066 - 0.003 - 0.015 + 0.000) = 19516

This value can be found in the third row of Table A.1, where the estimated income of such a household is given for the period 1991–2020.

Binary variables

Now consider the probability of the same household owning a PC in the fourth wave. Together with many of the topics of interest in the BHPS, PC ownership is a binary variable that is best modelled using logistic regression. The intercept, six main effects, wave effect and six interaction terms are added (multiplying the relevant coefficients by four in this case and noting that, with logistic regression, the reference category is not zero, but minus the sum of all the other coefficients relating to that variable), the resulting value exponentiated and added to one, and finally the inverse taken to produce the estimated probability of PC ownership:

PCOWN = 1 / (1 + EXP(- (-1.51 - 0.16 - 0.04 + 0.20 + 0.31 + 0.18 + 0.54 + 0.500 + 0.088 + 0.024 - 0.004 - 0.060 - 0.044 - 0.140))) = 0.47

This value can be found in the fourth row of Table A.1 where the estimated probability of such a household is given for the period 1991–2020.

The utility of this modelling approach is best demonstrated by changing the selected characteristics of the chosen household. Thus if, for example, this family lived not in Wales but in London, the probability would rise to 0.57; but, if that London household was in social rented accommodation, the probability would fall back to 0.43.

Using the regression model as a basis for forecasting

Using the above approach results in a methodology for projecting changes in variables of interest in a way that encompasses both inflation and distributional change. Taken together with the updating of the constraint variables, the three aspects of change previously identified have been covered. However, there is still the need to consider precisely how the attributes of the GHOSTs are to be updated in practice. At the moment, the household occupying a specific GHOST in, say, 2009 has the attributes of a real household interviewed in 1999. The six constraint variables for that household are known and the model shows how the *expected value* of a variable of interest for that type of household will have changed by that date. But how is the updating of the *actual* 1999 value to produce the 2009 *projection* to be done?

In the case of a continuous variable, a relatively simple procedure can be used. If a particular variable, in this case household income, is defined as:

AI99 to be the actual household income for the selected household in 1999;

EI99 to be the expected household income for the selected household in 1999;

EI09 to be the expected household income for the selected household in 2009;

PI09 to be the projected household income for the selected household in 2009;

then $PI09 = (EI09 / EI99) \times AI99$

Following the example of the Welsh household, then, if this specific household had an income of £15,000 in 1999, the projection for 2009 would be:

$$\begin{aligned} PI09 &= (EI09 / EI99) \times AI99 \\ &= (24029 / 21099) \times 15000 \\ &= 17083 \end{aligned}$$

By applying multipliers specific to the type of household under consideration, inflation and distributional change are combined while at the same time maintaining within-group variability, i.e. the variability that exists within households that share the same values on all six constraint variables.

In the case of a binary variable, the issue is somewhat more problematic. Here, the actual value for the selected household in 1999 can ony be zero or one; multiplication of the former will have no effect while multiplication of the latter produces a meaningless number. What is required is not a multiplier but rather a mechanism whereby sufficient of the zeros are converted to ones (or vice versa depending on the direction of the change). In the example Welsh household discussed previously, its expected probability of owning a PC in 1999 is 0.58, rising to 0.78 in 2009. If it is assumed that all those with a PC in 1999 have a PC in 2009 (bearing in mind that this is not an assumption regarding the behaviour of actual households but one required solely for updating our 2009 GHOSTs), then a conversion rate (CR) can be calculated, which can then be applied to those households without a PC in 1999:

$$(1 - 0.58) \times CR + 0.58 = 0.78$$

$$CR = 0.20 / 0.42$$

$$CR = 0.48$$

Forty-eight per cent of the households without a PC in the 2009 GHOSTs can be ‘given’ a PC in 2009.