Study of Past and Future Interregional Migration Trends and Patterns within European Union Countries In Search of a Generally Applicable Explanatory Model

Report on behalf of **Eurostat**
(Invitation to tender no: 2002/S 67-052015/EN; Lot 5)

Nicole van der Gaag, Leo van Wissen *
Phil Rees, John Stillwell, Marek Kupiszewski **

FINAL REPORT
DECEMBER 2003

*** Netherlands Interdisciplinary Demographic Institute**
(NIDI, P.O. Box 11650, 2502 AR The Hague, The Netherlands)

**** School of Geography, University of Leeds**
(SoG, Leeds LS2 9JT, United Kingdom)

Preface

Since 1985, the European Commission has commissioned four sets of internationally consistent population projections for the countries and the regions of the European Union (EU). The latest round of projections took place in the period 1995-1997, while the new revision has been planned for 2004. To improve the methodology behind these projections, the Commission also set up several background studies on each of the components of population change. The present report is the outcome of one of these studies, entitled 'Study on past and future interregional migration trends and patterns within EU countries – in search of a generally applicable explanatory model'. This study has been carried out as a joint effort of the Netherlands Interdisciplinary Demographic Institute (NIDI) and the School of Geography of the University of Leeds (SoG).

One of the activities of the study was an inventory of current practice and data availability in the countries of the European Union. We are very grateful to all respondents from the National Statistical Offices or other research institutes who kindly filled in the extensive questionnaires: Alexander Hanika (Austria), Leila Bellamammer, Micheline Lambrecht, Michel Poulain, André Doneux, Paul Willems (Belgium), Martin Stringfellow (England), Honkanen Ossi (Finland), Hansjörg Bucher (Germany), Marco Marsili (Italy), Hans Heida (the Netherlands), Maria Jose Carillho (Portugal), Claire Boag (Scotland), Margarita Cantalapiedra, Arlinda García Coll (Spain), Ake Nilsson, Sverker Lindbad, Mats Johansson (Sweden), Clive Lewis and Rhiannon Davies (Wales).

In June 2003, a workshop was organised at NIDI to discuss the preliminary results of the study with several internal migration experts. The results of this workshop have been included in the report as far as possible. We owe special thanks to the participants of the workshop for their valuable comments and advice: Hansjörg Bucher (Bundesamt für Bauwesen und Raumordnung, Germany), Harri Cruijsen and Evert van Imhoff (NIDI, the Netherlands), Arlinda García Coll (University of Barcelona, Spain), Frank Heins (Istituto di Ricerche sulla Popolazione e le Politiche Sociale, Italy), Mats Johansson (ITPS, Sweden), Petra Visser (The Netherland Institute for Spatial Research, the Netherlands) and Aarno Laihonen (Eurostat, Luxembourg), who also supervised the progress of the work for Eurostat.

Last but not least we would like to thank Dominic Brown, Oliver Duke-Williams and Daniel Vickers of SoG, and Dick Johan van der Harst of NIDI for giving assistance with data preparation and/or the cluster analysis.

The Hague, December 2003

Contents

1 Introduction

Since 1985, the European Commission has been involved in a programme for compiling internationally consistent population projections for the countries of the European Union (EU) on both the national and the regional (NUTS 2) levels. These basic projections are used for the preparation of European policies, regulations, directives and recommendations on various regional, economic and social issues. Four sets of population projections were made since 1985. The next revision of this set of projections is foreseen for 2004, when principal demographic data series collected in the population census round 2000/2001 are due to become available.

In between each set of projections, several background studies have been carried out, with the aim to improve the methodology for compiling internationally consistent projections at the regional (NUTS 2) level. The present study 'Study on past and future interregional migration trends and patterns within EU countries – in search of a generally applicable explanatory model', constitutes one of these background studies. The objective of this study is a cross-national analysis of contemporary interregional migration for regions at NUTS 2 level, and the development and improvement of methods to analyse, explain and project interregional migration trends and patterns. The point of departure of this study is the methodology used in the latest Eurostat scenarios, described in Van Imhoff *et al.* (1997), Van der Gaag *et al*. (1997a) and Van der Gaag *et al.* (2000), and named as the EUROPOP1995 model in the review by Rees *et al*. (2001).

1.1 General context

In general, the study addresses the following topics (linked to key questions):

1. Point of departure and aim of the study (or: what would we like to do?)
2. Current practice in the EU (or: what happens in the countries?)
3. Data issues and hypotheses (or: what is feasible, and what is not?)
4. Modelling internal migration (or: what did we do?)
5. How to make new scenarios (or: what should we do?)

In the remainder of the Introduction, each of these topics will be presented in more detail.

Point of departure and aim of the study

In the previous Eurostat regional projections, interregional migration flows are decomposed into a number of dimensions:

- overall time-dependent annual total migration rate, denoted T;
- regional relative out-migration rates O (relative with respect to overall rates); these region-specific rates were not time-invariant, and so are denoted as OT;
- regional in-migration shares D. These shares also vary with time: denoted as DT;
- an age-specific factor A, which is different for men and women S, denoted as an interaction term AS; these age-and sex-specific rates turned out to be origin- and destination-specific: ASO and ASD; and
- a distributional component, linking origins and destination, OD.

In short, the model decomposes the total migration flow matrix into an overall component, an origin- and a *destination*-component, and an age-sex component. In addition, there are interactions between various components. A major advantage of this approach is that it simplifies the structure of the matrix significantly. For scenario-making, only the time-varying dimensions need to be taken into account. It was shown that the OD interaction dimension was stable over time, as well as the region-specific age- and sex-profiles of out-migration and in-migration. Time-varying elements were:

- the overall migration rate T;
- the regional out-migration factor OT; and
- the regional in-migration factor DT.

These factors were used as parameters in the scenario model. This led to three types of questions:

1. How does the overall migration level vary over time and what should we assume for the different scenarios (high, baseline and low)?
2. How do the region-specific relative out-migration rates vary over time, and what should we assume in the three scenarios?
3. How do the region-specific relative in-migration shares vary over time, and what should we assume in the three scenarios?

The general philosophy underlying the scenarios was a distinction between convergence and divergence in trends. In the high scenario (which was more or less based on high economic growth), convergence in trends over regions was assumed. In the low scenario (low economic growth) divergence was the leading trend. In line with the economic assumptions behind the scenarios a high migration rate was assumed in the high scenario, and a low rate for the low scenario.

According to the evaluation made by Rees *et al*. (2001) of all four previous Eurostat scenarios, future projections should be based on the framework developed in the EUROPOP1995 model, but a number of possible improvements must be implemented. A major criticism of EUROPOP1995 was that country- and region-specific information was not taken into account sufficiently. The following recommendations are of relevance for the current study:

- the model of interregional migration based solely on statistical and demographic factors should be extended to incorporate models that include region-specific socio-economic determinants of migration for each life course stage; and
- consideration should be given in the fifth round of EU projections to moving from a "one size fits all" to "a best model for the country" approach when designing the interregional migration models.

These recommendations have been captured in the present study as far as possible. It has been examined to what extent the following new elements could be included in the model:

A life course perspective
Age was present in the 1995 model, but was fixed over regions and in time. In the present model, more attention has been given to the various stages in the life course, with different migration behaviour. The following stages have been used, along with the following age-categories:

	Life course group	Age group
-	children	0-14
-	students	15-19
-	young workers	20-29
-	family	30-44
-	old workers	45-59
-	elderly	60+

Each category may have a different migration pattern and this pattern may develop differently over time. Where possible, these age-categories have been applied to both sexes.

Non-demographic regional variables
In the previous scenario round in principle each region was treated in the same way. There was no explicit distinction between different regions, for instance urban and rural areas. This neglected a number of basic distinctions in the quality and characteristics of the regional system, and the population interactions within this regional system. In the present model, regional characteristics are explicitly taken into account in describing and explaining out-migration rates and destination patterns.

The resulting scenario model is able to produce results based on (1) purely demographic variables (in line with the previous EUROPOP1995 model) and (2) (socio-economic) explanatory variables. The purely demographic scenarios may be viewed as a trend-benchmark scenario, while the scenarios that are driven by explanatory variables include economic and other underlying assumptions. In principle, these scenarios can answer questions such as: what happens to migration and population if you reduce regional income inequality by half, or if you double regional income inequalities?

Current practice in the European Union

For the development of a new European sub-national migration model, the latest information and experiences with the models used in the countries themselves is a necessary prerequisite. This information has been collected through a survey sent to the National Statistical Offices in all EU countries with more than two NUTS 2 regions ((Kupiszewski and Kupiszewska, 2003). This survey updates two previous inventories: the Eurostat/NIDI inventory of regional projection model practice in EU countries (Van Imhoff *et al.,* 1994; Van der Gaag *et al*., 1997b), and the Council of Europe/Leeds inventory of internal migration information in 18 European countries (Rees and Kupiszewski, 1999a; 1999b). It is important to recognise that whilst international migration becomes interregional migration at a European scale, the focus of the current study is on migration taking place within each member state.

The survey provides an evaluation of the quantity and quality of the data that are available and an assessment of their strengths and weaknesses in forecasting at the regional level. In addition, the definitions of migration applied in each country were collected and analysed. This includes the temporal aspect of migration and administrative conditions needed to count a move as a migration.

Data issues and hypotheses

The EUROPOP1995 migration projections relied heavily on regional time series of migration data at the NUTS 2 level for the period 1990-1994. Since then these data have been updated with later years, and currently for many countries there are time series up to the end of the 1990s. However, data requirements for the suggested approach are high and not all data needed will be available for all countries. In the present study, internal migration has been modelled for four pilot countries: the Netherlands, Spain, Sweden, and the United Kingdom. As the study was planned to be based on Eurostat data, a first step was to evaluate demographic and socio-economic data series available at Eurostat.

Modelling internal migration

The aim of the study is to improve the methodology of the previous EUROPOP regional population projections. The core activity of the study, therefore, is the modelling exercise. Three countries have been analysed in detail: Sweden, the Netherlands and the United Kingdom. The strategy was to estimate models for the period 1991-1995, and to predict the flows in 1996-1998, using the estimated model coefficients for each of the countries. We modelled the internal migration process in two steps: first the out-migration rates, and second, conditional on out-migration, in-migration probabilities. Subsequently, Spain was used in order to see if the best common model for the other three countries was suitable for the Spanish case.

How to make new scenarios

In the final step of the study, a protocol has been formulated for devising interregional migration scenarios. In this protocol, it has been established to what extent the new methodology is applicable in all EU-countries. Preliminary results of the study have been discussed with internal migration experts of various countries. The protocol for interregional migration scenarios has been based on the results of the study and the discussions during this workshop.

1.2 Outline of the report

This report documents all activities within the study on past and future interregional migration trends and patterns within EU countries – in search of a generally applicable explanatory model. In chapter 2, a review has been made of recent theories and models explaining and projecting interregional migration behaviour. Details on the inventory and evaluation of existing sub-national population projections and internal migration data series for individual European Union countries can be found in chapter 3.

Chapter 4 focuses on the regional NUTS 2 classification of the four pilot countries (the Netherlands, Spain, Sweden and the United Kingdom) and contains an overview of all data used in the study. This refers to both internal migration data as well as the explanatory variables. In addition, tentative hypotheses are given on the relationship between internal migration and the explanatory variables.

The results of the modelling exercise will be given in chapter 5. In addition to the regional distribution of internal migration, i.e. regional out-migration and destination choice, we also discuss relationships between the overall time trend of internal migration intensities and various economic indicators.

To what extent the methodology tested is applicable in the other EU-countries is described in chapter 6. In the final section of this chapter, a protocol has been formulated for devising interregional migration scenarios. In chapter 7, finally, the most important findings are summarised, and recommendations and suggestions are given for further research.

2 Review and assessment of recent theories and models

2.1 Introduction

Population migration involves the relocation of individuals or households between geographical locations. It is a complex phenomenon not only because of the complexity of spatial patterns of movement that are involved but because of the myriad of motivations that influence the size and composition of flows between any two discrete areas and because of the imprecision of the data that are collected and used to analyse spatial patterns of flows over time. Migration can be measured in different ways on the basis of data obtained from alternative sources. However, whilst it is possible to utilise census, survey and registration data to provide some insights into directional patterns and temporal change nation-wide, it should be recognised at the outset that there is a conspicuous lack of data directly relating migration flows to the motivational factors that underpin the movements that take place and this has a limiting effect on explanatory analysis. Population censuses across Europe do not tend to ask people why they moved; registration data normally lacks a motivational dimension; and surveys that do ask questions of this type are usually localised and rarely comprehensive. As a consequence, those who seek to explain migration flows in a particular system of interest for a stated period of time are confronted with the task of trying to identify the determinants that are relevant and then to establish which of those explanatory variables are the most important. This is the challenge that has been taken up by many researchers and which has resulted in a plethora of studies involving different modelling approaches, measures of migration, and explanatory variables.

In contrast to the range of studies that have sought to *explain* internal migration using modelling methods, there is also another strand of pure and applied modelling work that has attempted to *project* internal migration based on current or historical trends and frequently in the context of the estimation and projection of sub-national populations. This tradition is an important part of regional demography and has its roots in the single region and multi-regional population projection approaches developed in the 1960s and 1970s. Whilst academic research on migration modelling has embraced both explanatory and projection methods, the application of migration models by national government departments or agencies has in the past tended to focus on the generation of internal migration projections from a migration sub-model independent of the main demographic projection system. The approach adopted in England during the 1980s is a prime example of this with the migration projections prepared by the Department of Environment being fed into the Office of Population Censuses and Surveys' population model (Armitage, 1986). In contrast, some countries have been attempting to link explanatory factors into their population projection systems since the 1980s. A good example is the Demographic Regional Economic Model (DREM) developed by the Central Statistics Bureau in Norway which attempts to take into account labour market factors determining migration flows between regions (Stambøl, 1991).

In this chapter of the report, these two approaches provide a broad framework for a review of migration models where emphasis is given to some of the most recent studies in this field. However, it is important to recognise some fundamental theoretical underpinnings to migration modelling work in general (Section 2) and to identify the range of factors that influence migration selectivity and which actually determine migration flows (Section 3). Explanatory migration models have been developed in a very wide range of spatial and temporal contexts and it is impossible to present a comprehensive review; Section 4 therefore focuses on the tradition of spatial interaction modelling that has its roots in the application of Newtonian gravitational principles in social science. Distinction is drawn between mathematical and statistical calibration methods of different forms of spatial interaction models and a more detailed summary is provided of a recent state-of-the-art two-stage migration model based on spatial interaction principles and calibrated using statistical regression. Subsequently, in Section 5, in the context of multi-state demographic models, attention is paid in particular to migration models developed and used for the projection of internal migration in countries of the European Union. Some conclusions from the review are drawn in the final section, that will provide certain guidelines for the model simulations that are reported later.

2.2 Theoretical underpinnings

A key distinction in migration modelling is that between micro and macro approaches (Stillwell and Congdon, 1991), a dichotomy which has parallels in economics and in psychology (White, 1980), for example.

Micro theory

Micro theory relates to the individual migrating unit (individual or household) and to the processes underlying the decision of the potential migrant to remain in the current location or to move somewhere else. It involves identification of those factors that influence this decision-making process: in the first instance, whether to stay or to move. Thereafter, it also takes into consideration the subsequent stage in the individual decision-making which involves choice between the alternative destinations that are available, once the decision to move rather than stay has been taken: whether to go to destination i or destination j. Since the choices at both stages are between discrete options (go or stay; go to i or j), the approach to migration modelling at the micro level is often known as the discrete choice approach (Maier and Weiss, 1991) and has its roots firmly fixed in the axiom of utility maximization since it is peoples' expectations about improving their own prospects in various locations that are at the heart of the decision-making process (Rothenberg, 1977).

The factors bearing on these decisions include both the characteristics of individual persons (such as age, marital status, household status) or wider family units (such as family size and structure) and the wider characteristics of the potential destinations (such as regional relativities of unemployment, wages or house prices). The relationship between migration behaviour and the changes that individuals experience as a consequence of progress through their life courses have been examined by various researchers since Thomas (1938), many in the context of intra-urban residential mobility and involving social psychologists such as Rossi (1955). Since utility is stochastic, a micro model formulation is likely to mean that the probability that an individual will choose destination region i is determined by an expression that compares the attributes of region i *vis a vis* those of the other possible destinations. The model which is calibrated empirically is a multi-nomial logit model. However, some potential destinations may be evaluated similarly because of preferences for certain types of area and this may lead to correlation in the random component of utility functions, leading to a contradiction of the assumptions of the model. Consequently, a number of studies, including Liaw and Ledent (1987), Hughes and McCormick (1989) and Van Wissen and Rima (1988) subdivide the decision-making process conceptually into two parts and consider that people's evaluations of alternative destinations are correlated. Nested logit models or multinomial probit models are used in these studies.

Macro theory

In contrast to micro theory, macro theory relates to aggregate migration flows and is more appropriate for setting migration in its labour or housing market context in order to deal with questions such as whether people migrate into areas where jobs are available or where prices are lower rather than the behavioural aspects surrounding the migration decision itself. Macro approaches are therefore concerned with investigating relationships between migration and objectively determined macro variables such as population sizes, unemployment rates, economic growth rates or environmental conditions. One traditional theoretical perspective in this context is that embraced by classical models of regional self-balance which suggested that migration is the equilibrating mechanism through which regions achieve adjustment, as, for example, when people move from regions with high unemployment to regions where unemployment is low. Myrdal (1957), however, argued that the selective nature of migration enhances regional differentials and therefore migration is disequilibrating rather than equilibrating. In contrast to the debate over the role of inter-regional migration in this context, the seminal contribution to migration theory offered by Lee (1966; 1969) in identifying push and pull factors influencing aggregate flows of internal migrants between regions has been of fundamental importance to those constructing macro models.

The distinction between micro and macro approaches thus provides a broad classification system for migration modelling and Cadwallader (1989) has articulated a valuable conceptual framework for understanding the relationship between the two approaches by suggesting that there are four sets of relationships:

- between aggregate migration and regional attributes that has been traditionally investigated by macro models;
- between the regional variables defined objectively and the subjective perceptions of those indicators by individual migrants;

- the integration of those perceptions about places into aggregate utility functions; and
- their subsequent translation into aggregate migration flows.

Data availability has been a major constraint on micro-behavioural modelling. There are relatively few national or regional surveys of migration motivation that provide spatial data on individual person, family of household decisions with regard to migration. In contrast, attempts to model the macro relationships between migration and factors deemed to be influential are more commonplace because of the availability of aggregate data on migration from censuses and registers and of explanatory variables from a variety of government and private sources. EUROSTAT's Regio database is an attempt to assemble a set of relatively consistent demographic and socio-economic variables for regions at different spatial scales across the European Union. The range of determinants in general is considered in the following section.

2.3 Selective influences and determinants of migration

An important distinction is between those characteristics of individuals or households that are indicative of higher or lower propensities to migrate and those factors that actually determine whether a move takes place and which destination is selected. Age is a typical example of the former. Age does not itself determine migration but people of different ages have very different migration propensities and migrate to different types of places because of the different motivations that influence their decision making. Job opportunities exemplify the latter; the opportunity to work elsewhere may well be the driving factor behind the migration of many in the labour force ages but may only be one of a combination of determinants. We can examine the range of selective influences on migration and the determinants of migration separately but should acknowledge that they are not mutually exclusive. A comprehensive review of determinants across a spectrum of dimensions is found in Champion *et al.* (1998) and the section below provides only a synopsis.

Selective influences

Demographic characteristics have a major influence on migration propensities. Age is a variable that changes for the individual in a regular and irreversible way over the life course, whilst sex is fixed at birth and persists. Migration intensities vary in a familiar way with age in most developed countries at different spatial scales as demonstrated by work at IIASA in the 1980s (e.g. Rogers and Castro, 1981). Migration rates tend to be high for young children, decline to school-leaving age (16 in the UK) and rise to a peak in the early twenties in response to higher educational and work opportunities. Thereafter, there is a decline throughout the child-bearing and child-rearing years which is parallel to the declining propensity of child migration. In some systems of interest, migration rates increase at the ages of retirement and may rise also at older ages (late 70s onwards) as more elderly people require family support or health service provision and migrate as a consequence.

The shape of the age-specific migration rate schedule reflects a number of life course transitions (e.g. leaving home, getting married, having children, retirement, *et cetera*) whose sequence and associated housing needs and distance of movement has been carefully documented by Warnes (1992). Moreover, the relationship between migration rates and age has been modelled by as a function of five components associated with childhood, employment, retirement, old age and a constant (Rogers *et al.,* 1978; Rogers and Castro, 1981; Rogers and Willekens, 1986). The model of migration intensity at exact age *a* for any zone has the form:

$$
\begin{aligned}
m^a = \; & b_1 \exp(-\alpha_1 a) \\
& + b_2 \exp\{-\alpha_2(a-\mu_2) - \exp(-\lambda_2(a-\mu_2))\} \\
& + b_3 \exp\{-\alpha_3(a-\mu_3) - \exp(-\lambda_3(a-\mu_3))\} \\
& + c
\end{aligned} \qquad (2.1)
$$

where the profile of the schedule is defined by seven parameters ($\alpha_1, \alpha_2, \mu_2, \lambda_2, \alpha_3, \mu_2, \lambda_3$) and the level of the schedule is determined by the remaining parameters (b_1, b_2, b_3, c). This model was operationalised in a general computer program called MODEL by Rogers and Planck (1984). It has been fairly widely used for smoothing erratic data and disaggregating from broad to narrow age bands and the methodology was used in the sub-national population projection model for England in the 1980s following a design by Bracken and Bates (1983) and Bates and Bracken (1987) as described by Boden *et al.* (1991).

Migration differentials between males and females are much less distinct than those between age groups but differences will be distinguishable when comparing migration intensities and patterns in certain contexts. Female rates may rise faster than males after age 16 to a slightly earlier peak than men and then declining at a rate slightly below men until retirement age. Thereafter, particularly in older old age, female rates may exceed those for males again. The gender differences in younger ages are partly a result of women leaving home earlier than men and marrying/cohabiting with men who are on average two years older, whilst in older age, higher migration amongst women is partly due to men dying earlier than their partners who may subsequently move. Differences in migration profiles are also evident between single, married, widowed and divorced groups (Devis, 1983).

Since Sjaastad's (1960) pioneering work on the human investment approach to migration analysis, our understanding of how age composition change influences migration has improved considerably. Plane (1992) explored the effects of demographic change on migration in the USA through an examination of migration rates of different age groups and cohorts over time, and the effect on total migration flows of the ageing of regional populations. Plane and Rogerson (1991) have borrowed Easterlin's (1980) relative cohort size hypothesis to explain migration levels. Babyboom generations, for example, experience more competitive conditions on entry into the labour market and hence fewer job opportunities tend to depress migration levels in comparison with smaller birth cohorts. Pandit (1997) has carried out a set of time series tests on the efficacy of the cohort size hypothesis *vis a vis* the business cycle hypothesis in the USA and tentative interpretations of the influence of birth cohort effects on net-migration for selected zones in the UK and Australia have been undertaken by Stillwell *et al.* (2001).

Whilst regular demographic influences have been shown to occur in migration intensities in different countries, fewer cross-national comparisons have been undertaken that have focused on the differences in migration propensities between various ethnic groups, social classes, those with different educational qualifications, those in different classifications of economic activity or those with different housing tenure characteristics. Various national studies have been carried out including that by Owen and Green (1992) in the UK, for example, that utilised the results of the Labour Force Survey to provide evidence of inter-regional migration differentials by ethnicity, by economic position and by educational qualification. Fielding (1992) has used data from the Longitudinal Study in the UK to demonstrate the extent to which inter-regional migration selects persons in higher-level occupations. Managerial and professional SOC groups show inter-regional migration rates that are between 50% and 90% higher than the average whereas craft, skilled manual and plant and machine operatives at the bottom end of the social spectrum have rates that are only 50% of the average. Another key selective influence on migration is housing tenure. Hughes and McCormick (1981; 1985; 1987) and Boyle (1993), for example, have argued that the management of public housing is responsible for discouraging longer distance migration of tenants in Britain, whilst those in owner-occupied housing are known to be more likely to move over longer distances than those in council housing.

This short review of selective influences serves to highlight the important issue that surrounds modelling based on data sets of aggregate migration flows. It is very likely that these bundles of individuals will conceal streams of selective migrants influenced because of their demographic, social and housing circumstances, but also motivated to move for a host of different reasons. Some of the key factors that actually determine migration are now presented.

Determinants

Gravity variables

Lee's classic study published in 1966 conceptualises migration as involving origins, destinations and the links between them. The characteristics of the origin may act as 'push' factors for potential out-migrants whilst the attributes of the destination reflect 'pull' factors that entice migrants to a particular destination. The separation of origins and destinations imposes a cost on migration and the term 'impedance' is often used to refer to the frictional effect of distance on migration. These factors were represented in the early formulations of the gravity model as gravity variables and were measured by the total populations of the origin and destination zones and the physical distance between them. Migration was considered to be a direct function of origin and destination size and an inverse function of distance. Thus, all other things being equal, places further apart therefore tend to have less to do with each other than places close together.

The size of the population at both origin and destination ends of the migration is the variable frequently used to standardise the migration measure and provide migration intensities that are comparable such as rates or velocities. Different measures of population size may be used, such as size of the labour force. Places which have larger populations in broad terms will tend to mean more attractions, services, opportunities and therefore be associated with higher levels of aggregate inter-regional migration. However, the scale of the regions used in the analysis may be important in this context, since as regions increase in size, a higher proportion of migrants will move intra-regionally. Moreover, there may be demographic structure characteristics of the aggregate populations which make certain places more attractive to particular age groups and population density provides an alternative specification of the mass variable, adjusting for area extent and also representing a measure of environmental conditions.

Considerable debate has centred on the measurement of distance in gravity models in particular since it may be argued that physical distance does not reflect either the social costs of moving or time costs that may not be proportional to distance. Moreover, measures of physical distance can vary from Euclidean straight-line distances between zone centroids to road mileage distances or network-weighted distances calculated on the basis of shortest surface so as to take account of the effect of estuaries. Areas that share a boundary tend to have more migration between them, because this migration will include a proportion of short-distance moves from one side of the boundary to the other. Consequently, several studies have used contiguity variables that take the value of 1 for zones that share boundaries with each other, and 0 for other pairs of zones. Finally, it is apparent that areas situated in high population-density regions are likely to be less attractive as destinations to migrants, everything else being equal, because of increased spatial competition between destinations (Fotheringham, 1986). A destination accessibility or potential variable for zone i (POT_i) which measures the degree of spatial competition faced by a destination zone from nearby destination zones may be created as:

$$POT_i = \Sigma_{j\, j \neq i} (P_j / d_{ij}) \qquad (2.2)$$

where P_j is the population of zone j and d_{ij} is the distance between each pair of zones.

This competing destinations effect can be extended by incorporating a set of variables that describe each zone surrounding a destination. The following formula might be used to compute a so-called regional variable associated with one zone i, REG_i:

$$REG_i = [\Sigma_{j\, j \neq i} (X_j/X_i)\, d_{ij}^{\beta}] / \Sigma_{j\, j \neq i}\, d_{ij}^{\beta} \qquad (2.3)$$

where X_i is the value of variable X for zone i and X_j represents the value of X at one of the other zones in the system. The formula produces a distance-weighted average ratio of X_j to X_i where nearby zones are weighted more heavily in the calculation than more distant ones. The value of the distance decay parameter β (e.g. –2) gives a reasonably differentiated surface of *REG* values. Values of β less negative than this give a surface which is smoother; values of β that are more negative will give a spikier surface. Values of $REG_i > 1$ indicate that X_i is generally smaller than its neighbours. Values of $REG_i = 1$ indicate that X_i is generally very similar to its neighbours. Values of $REG_i < 1$ indicate that X_i is generally larger than its neighbours. The X variables would typically be the sort of explanatory variables identified in the following sub-sections which are subdivided into economic, labour market, housing market, environment and policy variables.

Economic variables
Longer distance migrants tend to have a higher probability of changing their place of work as well as their place of usual residence when they migrate. These migrants, together with their partners or families in some cases, are likely to be influenced by relative regional economic prosperity. In the UK, young aspiring business executives are attracted to the South East (and to London in particular) because this is a dynamic region where economic growth is relatively buoyant and where company development appears to be successful; in the same way, dynamic provincial centres will attract migrants within their own regions. Students are attracted to dynamic places where there are lots of activities taking place, as well as opportunities to get part-time jobs. Consequently, variables that measure levels of prosperity, such as gross domestic product per capita, or the number of new business registrations, together with those that identify how conditions are changing over time, are likely to be important influences, though these may be 'picked up' by variables that characterize the labour market.

Labour market variables
On many occasions, levels of prosperity are reflected in the conditions of the job market. Labour market factors are seen as potentially important both in prompting out-migration from an area as well as in influencing people's destination choices. Like GDP per capita, they would normally be interpreted as measures of the overall economic environment of an area, and include measures such as the levels of employment and changes in jobs as well as unemployment rates and changing unemployment conditions. A number of labour market variables, particularly those related to occupation and income, such as the proportion of the workforce employed in agriculture or the magnitude of skill shortages in particular sectors or the relative wage or salary rates, can also serve as indicators of the economic performance of areas, as can measures of the tightness of the housing market.

Housing market variables
Housing factors form a critical element underlying migration patterns, but they have complex interactions with migration and need especially careful treatment. On the one hand, some housing measures such as high house prices and low vacancy rates can reflect the strong economic performance of an area and indeed of neighbouring areas within commuting distance. On the other hand, these factors can directly influence the opportunities for in-migration, with high house prices acting as a deterrent and high vacancy rates as an attraction. Even here, however, the effect may not be as simple as this, since high house prices may lead to more rapid in-migration in anticipation of future price rises and very high vacancy rates may serve to undermine confidence in an area and deter people from moving there. The size, composition and quality of the housing stock can also influence both the level and the type of migration. Most obviously, the number of new houses constructed or the number of housing demolitions are likely to be very important determinants of migration. Housing tenure, besides being a reflection of the social composition of an area, is also known to affect migration patterns, most notably the well-documented problems that people moving between local authority areas have in accessing council housing.

Environment variables
In the current era within advanced economies, environmental factors play a major role in people's residential moves, both in prompting exits from areas and in acting as 'pull' factors. The term is used here in its broadest sense, covering all the physical, economic, social and political aspects that affect both the everyday quality of life and the longer-term trends in life chances. This category can therefore be considered to include most of the factors mentioned under the other headings above, insofar as they bear upon the overall quality of an area and of the neighbourhoods that it comprises.

Among those not specifically considered are variables relating to derelict and vacant land, variables relating to the pattern of development, such as the proportion of new housing on brownfield land, variables such as population density, settlement size and level of urbanization, variables relating to crime and anti-social behaviour, variables measuring climate and air quality, and variables relating to sports and leisure activities. Also included under the environment theme might be variables relating to physical attractiveness, such as accessibility to scenic areas, number of listed buildings, number of visitors, the extent to which an area acts as a dormitory for commuters to relatively distant jobs, the accessibility to international air passenger connections. A 'bright lights' variable that measures access to theatres and concert halls might also be considered as important for certain migrant groups.

Policy variables
Public policy variables relevant to migration behaviour include not only direct interventions such as migration incentives and migration policy (such as the distribution of asylum seekers from reception centres to allocated dwelling spaces) but also indirect influences through the uneven effects of government grants, local taxes, defence spending, higher education expansion and the amount and location of land approved for house building that is part of the physical planning process.

In general, it may be more satisfactory to estimate the role of public policy, past or anticipated, by reference to variables representing the aspects that public policy seeks to alter. For instance, the migration impact of a regional development initiative can be assessed by reference to, for instance, the number of extra jobs, while the impact of a policy that alters the availability of land for house-building can be studied via changes to the number of housing completions. Finally, in contrast to public policy, many organisations and private companies have staff recruitment and mobility policies (both internal and international) that result in inter-regional migration and for which it is very difficult to obtain any detailed information.

The review we have provided demonstrates the complexity that surrounds the phenomenon of inter-regional migration. This is accentuated by the fact that on many occasions, people tend to move and to choose their destinations on the basis of a unique combination of reasons. In concluding this section, it is also essential to recognise that different variables will be required to explain or project migration at different levels of aggregation. Thus, for example, trends over time in the overall migration intensities for one country are likely to be associated with fluctuating economic conditions, interest rates or mortgage rates, whilst regional out-migration rates for one period of time may be associated with regional prosperity indicators relating to each region and region indicators relative to the national average. On the other hand, the explanation of origin-destination migration flows between any two regions would need to embrace not only the characteristics of the two regions concerned and their distance apart, but also their characteristics *vis a vis* the characteristics of the regions that make up the rest of the system. These distinctions are taken up in the next section in which we consider alternative modelling approaches.

2.4 Migration models: explanatory approaches

Non-demographic models use additional non-demographic information for explaining and predicting migration patterns. There are many different forms that could be classified under this umbrella and which might be applied to explain regional differences in out-migration, the relative attractiveness of destination regions for in-migration, the net balance between out-migration and in-migration across a set of regions, or the spatial distribution of migrant flows between origins and destinations. The initial focus of the review that follows in this section is on the latter, and more specifically on the family of spatial interaction models that has a long history of application in geography and transportation sciences.

Early gravity models

Ravenstein recognised the importance of the frictional effect of distance on migration in formulating his laws of migration back in the nineteenth century (Ravenstein, 1885) but migration models based on gravitational features were first developed in the 1940s (Zipf, 1946). These models incorporated terms measuring the masses of each origin and destination and of the distance between them and were calibrated statistically using log-linear regression techniques. Modifications were made to these early Newtonian gravity models by introducing parameters to weight the influence of the origin and destination factors and by experimenting with alternative distance functions.

Spatial interaction models: mathematical formulations

One of the shortcomings of these early approaches was the inability of the OLS regression formulation to predict interaction that was consistent with observed flows from each origin and to each destination. This was remedied by the introduction of so-called balancing factors (Wilson, 1967) to ensure internal consistency within the model and the derivation of the same model based on entropy-maximising techniques (Wilson, 1970). Wilson's family of four models of spatial interaction between any two zones i and j took the following general form:

M_{ij} = Scaling factor (or balancing factors)
* Origin out-migration (or repulsive factor)
* Destination in-migration (or attractiveness factor)
* Distance function (with distance decay parameter) (2.4)

where a scaling factor was used when no observed out-migration of in-migration totals were known so that the sum of all the flows predicted in the origin-destination matrix was constrained to the total number of migrations observed in the system (the so-called unconstrained case). Attractiveness factors were used as proxies for mass terms when out-migration or in-migration totals were unknown. When out-migration or in-migration totals were available, balancing factors replaced the scaling factor to ensure that the row or column elements of the predicted matrix were consistent with the observations. The doubly constrained model of migration between regions *i* and *j* incorporated balancing factors for both origins and destinations (A_iB_j), mass terms (O_iD_j) and the distance function (d_{ij}) used was typically either a power function (as shown below) or an exponential function $\exp(-\beta d_{ij})$:

$$M_{ij} = A_i B_j O_i D_j d_{ij}^{-\beta} \qquad (2.5)$$

These mathematical models were calibrated using an automatic Newton Raphson search routine that generated an optimum distance decay parameter by iteration from a given starting value on the basis of a measure such as the convergence between the predicted and observed mean migration distance (Stillwell, 1984; 1991). This approach was extended with the calibration of zone-specific distance decay parameters by Stillwell (1978) and the incorporation of a competing destinations variable to remove the effect of spatial structure by Fotheringham (1983; 1991). More recently, Fotheringham *et al.* (2001) have shown how the competing destinations spatial interaction model makes explicit the linkage between spatial choice behaviour at different levels in the spatial hierarchy.

Whilst models of this type were used typically to estimate missing information in a historical context, less commonplace are examples of the application of these types of spatial interaction models for migration projection. One example is the model developed by Rees *et al.* (1990) to project ward populations in Swansea. In the context of projection, independent projections are required of out-migration and in-migration that may be derived from the extrapolation of historical trends or may be connected with projected explanatory variables as summarised by Stillwell (1991). Examples of studies in which projections of migration were tested against observed data are almost non-existent.

Spatial interaction models – statistical formulations

In parallel to the development of mathematically calibrated spatial interaction models, statistical modelling of inter-regional migration has also evolved and new forms of models have been introduced from the baseline gravity model specification outlined by Congdon (1991) in which the variables were log transformed and which has the form:

$$\log (M_{ij}) = b_0 + b_1 \log (P_i) + b_2 \log (P_j) + b_3 \log(d_{ij}) + \varepsilon_{ij} \qquad (2.6)$$

where b_0 is the constant and b_1, b_2 and b_3 are the regression coefficients associated with the relevant population terms and distance, and where ε_{ij} is the random error term associated with each interaction. This general linear model formulation is equivalent to an unconstrained spatial interaction model in the Wilson family.

One of the earliest extensions was that of Lowry (1966) who extended the set of independent variables on the right-hand side of the equation and there have been a large number of studies subsequently that have sought to identify the most important determinants of migration. One of the basic assumptions of the linear model is that the observations are independent of one another and that the relationship between migration and the predictor variables is the same across each zone in the system of interest. The recognition that there are likely to be local variations in parameters has led to the application of geographically weighted regression (GWR) (Fotheringham *et al*., 2002) and the re-specification of the model in the following form:

$$\log M_{ij}(g) = b_0(g) + b_1(g)\log(X_i) + b_2(g)\log(Y_j) + \varepsilon_{ij} \qquad (2.7)$$

where X_i and Y_j represent explanatory variables, and (g) indicates that the parameters are to be estimated at a location whose co-ordinates are given by the vector g.

The restrictive assumptions associated with the log-normal model have also led to the emergence of new statistical models based on the Poisson distribution (Congdon, 1991; Flowerdew, 1991). In the log-normal model, the error term and hence the dependent variable are assumed to be log-normally distributed continuous variates and the variance of the errors is constant regardless of the size of the estimation flow. The issue here is that the migration dependent variable is likely to be measured in discrete units (integer counts of persons) and follows a discrete probability distribution. This is also particularly important when there is likely to be a large number of small flows in the origin-destination matrix and a much smaller number of small flows. In terms of model structure, this means that the Poisson regression equation becomes:

$$M_{ij} = \exp(b_0 + b_1\log P_i + b_2\log P_j + \log d_{ij}) + \varepsilon_{ij} \qquad (2.8)$$

In generalised linear modelling, a likelihood ratio statistic is used to assess how well the model fits the data. This statistic is called the deviance, D, and is calculated in Poisson regression as:

$$D = 2\,(\Sigma_{ij}\,\mathrm{Obs}M_{ij}\log(\mathrm{Obs}M_{ij} / \mathrm{Pred}M_{ij})) \qquad (2.9)$$

As the number of flows and the size of the flows increases, the deviance converges to the chi-squared distribution. Thus, the size of the deviance can be used to assess the goodness-of-fit of the model. Scholten and Van Wissen (1985) compared the performance of spatial interaction models with log-linear approaches and concluded that using log-linear models with historical interaction parameters performed better than other approaches in terms of model fit and prediction. Flowerdew (1991) demonstrated that the possibilities of fitting Poisson regression models on quite large data sets using GLIM and the Poisson regression approach has been developed and adopted in several studies since then including Flowerdew and Lovett (1988), Amrhein and Flowerdew (1992), Bohara and Krieg (1996) and Boyle *et al.* (1998). More recently, the application of origin-specific Poisson models calibrated using GWR has been undertaken by Nakaya (2001) and similar models have been used to compare interregional migration in Japan and Britain by Yano *et al.*

(2003). We consider the Poisson approach further in the next section where we present a more detailed case study of recent migration modelling in the UK and also in chapter 5 where we report on its use in demographic models.

Two-stage migration modelling

As indicated earlier, there is evidence that individuals often conceive of their migration as a two-stage process with worsening conditions at an origin eventually reaching a threshold level at which they decide to leave and then conditions at various locations being examined in order to decide on a suitable destination. This principle underpins the development of two-stage migration models, for which a state-of-the-art example is found in a research project completed by a team based at the Universities of Newcastle-upon-Tyne and Leeds. This study was undertaken for the Office of the Deputy Prime Minister (ODPM), formerly the Department of Transport, Local Government and the Regions (DTLR) and involved the calibration of a policy-sensitive model of internal migration in the UK and the development of a user-friendly planning support system known as MIGMOD (MIGration MODeller). Detailed accounts of the work are provided in ODPM (2002), Champion *et al.* (2002), Fotheringham *et al*. (2004), and Rees *et al.* (2004).

MIGMOD approach

The central features of this approach are the separate modelling of: (a) out-migration from each area based on a set of determinant variables (Stage 1); and (b) the distribution of migrants between destinations also based on a set of determinants (Stage 2). The project also involved the development of an operational, user-friendly combination of Stages 1 and 2, enabling the model user to quickly set up and run a range of 'what if?' scenarios, to view the large volume of inputs and outputs, and to develop a selection of scenarios of determinant variables reflecting desired policy options.

The data used in the model was from a time series of movement events recorded when National Health Service (NHS) patients re-register with doctors in different Family Health Service Authority (FHSA) areas across the UK. These data are collected in a central register (NHSCR) and provide a consistent time series of data on both zonal out-migration and origin-destination migration from 1983-84 to 1997-98. The data have been compared with 1991 Census migration by Stillwell *et al.* (1995) and used to monitor inter-censal migration patterns (Stillwell, 1994). The spatial system which defines the migration is a set of 98 FHSAs that are coincident with shire counties (England and Wales), metropolitan districts (England) and groups of London boroughs, together with Scotland and Northern Ireland as single regions. Whilst the availability of fifteen years of annual out-migration data allowed some time effects to enter the Stage 1 modelling, consistent origin-destination matrices of migration flows were only available for seven time periods which was not sufficient to allow time-related variables to enter the Stage 2 model. This model structure conforms to the consistency principles for spatial interaction models first defined by Wilson (1967; 1972).

Given the arguments spelt out earlier in this review about patterns of migration and reasons for moving varying significantly with a person's position in the life course, it was considered essential to disaggregate the model by age and by sex. Seven age groups were eventually chosen for the model corresponding to childhood/schooling ages (0-15), the ages at which adolescents leave home for higher education (16-19), the ages at which students leave higher education for their working and partnership careers (20-24), the ages when they look for career advancement (25-29), the family formation ages (30-44), the later working ages (quiescent in terms of migration)(45-59), and the retirement and older ages (60+).

Thus, the state-space involved calibrating the two-stage migration model for seven age groups and two sexes. The options of calibrating the Stage 1 model for each origin or for clusters of origins were ruled out in favour of an 'all origins together' calibration. Consequently, fourteen separate models were calibrated. However, the situation for the Stage 2 destination choice model was different and an origin-specific distribution model was adopted, allowing the determinants to have different influences on the outcomes for each origin. For Stage 2, therefore, it was necessary to calibrate 98×14 or 1,372 separate models (for just one year).

The assembly of the data needed for calibrating the Stage 1 and Stage 2 models proved to be a major task, both for the migration flows used as dependent variables and for the set of determinants used as independent variables. Data were obtained for 139 potential determinants of out-migration and 69 potential determinants of migration destination choice. Some explanatory variables were cross-sectional; others were available as a time series; some variables were lagged. In addition to variables measuring the characteristics of each zone, national variables were included and also regional variables were calculated for the Stage 1 model that were designed to capture the possible pull effects on out-migration caused by conditions elsewhere in the country.

Stage 1: Out-migration model
The volume of out-migration from an origin zone i is predicted as:

$$O_{it}^{m} = omr_{it}^{m}\, P_{it}^{m} \qquad (2.10)$$

where O_{it}^{m} is the total out-migration of migrant group *m* from zone *i* in time interval *t*, omr_{it}^{m} is the out-migration rate from origin *i* in time unit *t* for migrant group *m* (one of the 14 age-sex groups) and P_{it}^{m} is the population of migrant group *m* at risk of migrating from origin *i* during time interval *t*. The general form of the out-migration rate is as follows:

$$omr_{it}^{m} = f(X_{it/t-1}^{m}, Y_{it/t-1}^{m}, Z_{t-1}) \qquad (2.11)$$

where omr_{it}^{m} is the out-migration rate for migrant group *m* from zone *i* in time interval *t*, $X_{it/t-1}^{m}$ is a vector of origin attributes in either year *t* or *t-1* (lagged by one year); $Y_{it/t-1}^{m}$ is a vector of distance-weighted attributes describing the situation in other areas in either year *t* or *t-1* and Z_{t-1} is a vector of attributes describing the national economic situation as it affects the overall volume of migration in year *t-1* (lagged by one year).

Much debate focused on which specific form the model should take. Multiplicative or additive? Logged (the multiplicative option) or unlogged variables? Should non-linear forms of the variables such as quadratic forms be considered? Initially a multiplicative model was calibrated with the specific form:

$$omr_{it}^{m} = \exp(K^{m})\, A^{m}\, V_{1}^{a1m} \ldots V_{53}^{a53m} \exp(a_{54}^{m} V_{54}) \ldots \exp(a_{58}^{m} V_{58}) \qquad (2.12)$$

where the term $\exp(K^{m})$ is the intercept in the log-log regression, specific to each age and sex group. An age and sex-specific adjustment factor A^{m} is required because of a statistical bias in the intercept estimate in a log-log regression. Most of the explanatory variables (V_{1i} to V_{53i}) became log functions in the log-log regression, although some (V_{54i} to V_{58i}) could not be logged due to negative values and were represented in the exponential form. The regression parameters, represented by power coefficients, a_{v}^{m} are specific to each variable v, and to each migrant group m.

The initial specification of the model had a number of drawbacks. A number of investigations were undertaken to identify better model specifications:

- There was strong multicollinearity between sets of determinant variables; principal component analysis was used to produce a much-restricted set.
- The combination of power and exponential relationships was awkward; therefore a straightforward linear model was used.
- The assumption that a monotonic relationship existed between a dependent variable and determinant variables may not be correct; therefore the quadratic form of each variable was introduced as a potential determinant.
- There was tendency for certain flows, particularly between local government unit's intern London, to be underpredicted; dummy variables were used to correct for this.
- Initially, no allowance was made for time trends in migration; in the revised model a time trend was used.
- In the first model, variable selection was arbitrary; in the revised model, stepwise regression was used based on significance of parameters.

The final form of Stage 1 model related the adjusted out-migration rate for migrant group m in zone i at time t to a new series of independent variables comprising cross-sectional (X), regional (Y) and national (Z) indicators. The linear model of out-migration rates has the basic form:

$$omr_{it}^{m} = \kappa^{m} + \sum_{p} \alpha_{p}^{m} X_{pit}^{m} + \sum_{q} \beta_{q}^{m} Y_{qit}^{m} + \sum_{r} \gamma_{r}^{m} Z_{r}^{m} + \varepsilon_{it}^{m} \qquad (2.13)$$

To this were added quadratic terms, a linear time trend and a dummy (LD) for London:

$$\begin{aligned} omr_{it}^{m} = {} & \kappa^{m} + \sum_{p}\alpha_{p}^{m} X_{pit/t-1}^{m} + \sum_{q}\beta_{q}^{m} Y_{qit/t-1}^{m} + \sum_{r}\gamma_{r}^{m} Z_{rt-1}^{m} \\ & + \sum_{p}\delta_{p}^{m}(X_{pit/t-1}^{m})^{2} + \sum_{q}\eta_{q}^{m}(Y_{qit/t-1}^{m})^{2} + \sum_{r}\theta_{r}^{m}(Z_{rt-1}^{m})^{2} \\ & + \psi^{m} T_{t} + \zeta^{m} LD_{i} + \varepsilon_{it}^{m} \end{aligned} \qquad (2.14)$$

where ε_{it}^{m} is the error term for each zone, time and migrant group combination.

Stage 2: Destination choice model

Stage 2 involves the calibration of a migration destination model that distributes the total number of out-migrants from zone i to each of the destination zones based on the characteristics of each destination zone and the separation between the origin and each destination. It is a spatial choice model of the destinations chosen by migrants from an origin. The model to be calibrated has the general form:

$$M_{ij}^{m} = O_{i}^{m} \prod_{p} X_{pj}^{\alpha pim} d_{ij}^{\beta im} \ / \sum_{j} \prod_{p} X_{pj}^{\alpha pim} d_{ij}^{\beta im} \qquad (2.15)$$

where O_i^m is the volume of out-migration of type m from origin zone i; X_{pj} is an attribute of zone j that affects the choice of j by migrants from i; and d_{ij} is the distance between i and j. The X variables are raised to powers, α_{pim}, specific to each variable p, origin i and migrant group m, while the distance variable is raised to the power β_{im} specific to each origin i and migrant group m. The parameters of this model indicate the sensitivity of migration flows to particular destination characteristics; they indicate what features of a destination make it attractive to migrants and which features make it unattractive. For example, a relatively large score on an attribute with a positive parameter estimate would make a destination attractive to migrants, *ceteris paribus*, while a relatively large score on an attribute with a negative parameter estimate would make a destination unattractive to migrants, other things being equal. Models of this type have been discussed earlier in the chapter and have a long history in the analysis of spatial interaction patterns.

The model is calibrated separately for each of the origins and each of the 14 migrant groups. For any origin, $O_i^m / \sum_j \prod_p X_{pj}^{\alpha pim} \ d_{ij}^{\beta im}$ will be a constant (k_i^m) so that the origin-specific model is then simply,

$$M_{ij}^{m} = k_{i}^{m} \prod_{p} X_{pj}^{\alpha pim} d_{ij}^{\beta im} \qquad (2.16)$$

The initial calibration of this model was by OLS regression by taking logs of both sides of the equation to make it linear-in-parameters. However, the count of the dependent variable is a count of migrants and therefore is likely to be Poisson distributed rather than normally distributed. This means that one of the key assumptions of OLS is probably not met. Poisson regression was therefore preferred since it assumes the conditional mean of the migrant variable has a Poisson distribution and avoids the need for making some approximation to zero flows. An initial set of 69 explanatory variables was reduced to 27 following a qualitative assessment of each of the variables in the data set and an examination of multicollinearity amongst the independent variables.

This approach to migration modelling provides an example of how the form and content of a model may evolve over the duration of a research project. The model was constructed in two phases and, as indicated above, many changes were made to the original specifications to overcome difficulties. The prediction of out-migration flows from origins and their distribution to destinations were combined in Stage 3 into an operational model for the sponsoring organisation so that scenarios related to policy could be implemented.

2.5 Migration models: demographic approaches

The importance of internal migration as a component of population change has been widely recognised by those responsible for creating sub-national population estimates and projections. Consequently, a second genre of approaches to modelling migration has emerged within the field of multi-state demography whose aim has been to generate projections of migration flows without involving the type of detailed explanatory factors discussed in Section 2. Wilson (2001) provides a detailed review of the evolution of multi-regional demography, with a clear specification of the model equations. The following synopsis draws on this review.

Multi-state population projection modelling

The earliest population projections were usually produced using a cohort component model which, in the case of a single region, involved the estimation of the population at the beginning of a projection period, the projection of the number of births during the future time period and the survival of those in existence or being born during the period. Early examples of uni-regional models include those developed by Bowley (1924) in Britain, Weibol in the Netherlands (de Gans, 1999) and Whelpton (1936) in the USA. Leslie (1945; 1948) re-wrote the uni-regional model in matrix notation whilst others (e.g. Plane and Rogerson, 1994) demonstrated how the model could be expanded to include net-migration either in the form of flows or rates.

As far as modelling the migration component was concerned, it was the development of multi-regional demography in the mid-1960s that heralded the proper specification of inter-zonal flows rather than net-migration balances in projection models. Andrei Rogers (1966; 1967; 1968) pioneered the development of the Leslie matrix for a multi-region system and the creation of multi-region life tables (Rogers, 1973). He also provided the theoretical rationale for the use of migration flows rather than net balances in Rogers (1990). An alternative approach to the Rogers' multi-regional survival model known as accounts-based modelling was developed during the 1970s by Rees and Wilson (1973; 1975; 1977), Wilson and Rees (1974) and Rees (1981). Rees and Wilson constructed accounts-based models for transition data (involving the migration of those in existence at one point in time who were living at another address at an earlier point in time) in the first instance before applying similar techniques to movement migration (counts of moves taking place in a period irrespective of existence at the beginning or end points) (Rees, 1984). Willekens and Drewe (1984) brought the Rogers and Rees approaches together by switching from a dependence in the model on the multi-regional life table to period-cohort rates.
Thus, demographic models have developed from models requiring little information about migration to models requiring maximum information about migration, i.e., from aggregate net-migration balances, through migration pool, to migration flow information disaggregated by single year of age and sex. Population projection modelling has become more sophisticated as the migration component has been specified with more precision. Within this demographic modelling context, there are two key questions that relate to the internal migration component. The first of these is how to incorporate some form of change into the parameters that govern the intensity and pattern of migration during the projection period. The second is how to deal with the problem of huge data arrays when the origin-destination-time-age-sex dimensions are cross-classified. We

discuss briefly each of these issues in turn, before reviewing some of the more recent research undertaken in the context of the development of multi-state models for European NUTS regions.

Temporal variability in model parameters

Many multi-regional population projection models do not in fact include any temporal variability in the origin-destination migration intensities upon which the model is based; they adopt the Markovian assumption that migration intensities will not change from one period to the next. However, there are some approaches that do try to build in some temporal variance. Plane and Rogerson (1986) discuss the use of causative matrices of ratios which link matrices of Markov intensities from one time period to another in the same way that it is possible to extrapolate from a geometric regression based on two data points. Feeney (1973), on the other hand, adjusts the Markov migration intensity by allowing the distribution of out-migrants to vary over time. This works by adjusting the base period intensity using the ratio of the destination region's share of the national population (excluding the origin region) at the start of the projection period to the same share recorded in the base period. The model is written as a probability of migrating between an origin and a destination in a projection period. An alternative probability approach is that termed the destination-population-weighted (DPW) model (Plane, 1982) which incorporates a balancing term to ensure that the probabilities sum to one. Some authors, including Fielding (1992) and Courgeau (1995) have suggested defining an origin-destination migration intensity based on the populations of both the origin and destination. This measure of migration velocity can be used in the same way as the traditional Markovian intensities but would require an adjustment if temporal variation was required. Pioneering work on the temporal stability of migration was undertaken in the 1980s in the Netherlands by Baydar (1983) who decomposed migration flows into an overall component or the total number of migrants in year t (N_t), a generation component or the probability of out-migration from region i in year t (o_{it}), and a distribution component or the probability of in-migrating to region j given origin i (p_{ijt}):

$$M_{ijt} = N_t \, o_{it} \, p_{ijt}, \quad i \neq j \tag{2.17}$$

and used a log-linear model to calibrate the parameters which quantify the time dependence of the different variables and thus identified the most stable and volatile components.

Shrinking dimensionality

The second issue revolves around the necessity to shrink large dimensional multi-regional models since the modern form of a demographic sub-national migration model is the multi-state model that uses migration flow information by age, sex, region of out-migration and region of in-migration. In its pure form, the multi-state migration model is highly descriptive: it has a separate parameter for every piece of information of the migration pattern. This means that the data requirements for the full multi-dimensional model are very large indeed. The creation of population projections for a system of 30 regions with 100 age groups and two sexes in any one year would involve 30(origins) x 29(destinations) x 100(age groups) x 2(sexes) = 174,000 flows. Research by van Imhoff *et al.*

(1997) has shown how far it is possible to simplify (shrink) the structure of the multi-regional model before the resulting loss of information and accuracy becomes unacceptable. Their approach is reviewed in the following section.

Poisson modelling in a multi-state projection context

From a methodological point of view the multi-state model can be viewed as an accounting structure for a spatial interaction model. Both developments have converged using the framework of the Poisson regression model. The approach by van Imhoff *et al.* is particularly relevant here since their study was conducted in the context of the development of regional population projections at NUTS 2 level across the European Union and is therefore highly appropriate as a forerunner of the current project. Moreover, the projection method parallels that of the MIGMOD approach by separating the modelling into two stages: (i) the projection of out-migration by age and sex from each region and (ii) the allocation of this pool of out-migrants to destinations. The second stage is known as a 'migrant pool' model because in-migration to destinations depends only on the size of the pool and not on the composition of the pool by region of origin. In the framework of log-linear modelling, the pool model corresponds to a hypothesis of independence between the origin and destination.

The approach assumes that interregional migration is classified along five dimensions referred to by letters: O (representing region of origin); D (region of destination); A (age); S (sex); and T (time period). Consequently, the observed count of migrants (or moves when registration data are being used) is represented by M_{ijast} where i and j are particular regions, a refers to one age group, s refers to males or females and t refers to one time period. The objective is to develop a model that describes each migration flow (or its corresponding rate) as the product of a limited number of parameters and then to examine the relative significance of the parameters. This approach therefore seeks to answer questions such as: Are the parameters representing sex more important than those representing age? How important is the origin effect? Is the time trend significant? It also allows the significance of relationships between dimensions to be identified, the so-called interaction effects, e.g. between particular origins and destination regions or between certain age groups and sex.

The Generalised Linear Modelling framework provides a suitable context for estimating the parameters of this type of model and log-linear regression models can be calibrated using a maximum likelihood algorithm available in the GLIM software package. The Poisson model is particularly useful because it produces unbiased parameter estimates, even in the case of over-dispersion in the data set (Davies and Guy, 1987). The parameter values are automatically normalised in GLIM and there is always a one-to-one correspondence between the number of parameters and the degrees of freedom in any model. Unlike the MIGMOD approach in which separate log-linear models are fitted for each age group for males and females, log-linear modelling using GLIM in this context can make use of the complete data sets and therefore it can take a long time to calibrate all the parameters. Once calibrated, the goodness of fit of a model is measured in GLIM using a measure of deviance defined in equation (2.9). The deviance statistic can be compared with a chi-square distribution with degrees of freedom equal to the number of degrees of freedom in the model. In other words, as the number of flows and the size of the flows increases, it converges to a chi-squared distribution. Thus, if the deviance is greater than the critical chi-squared value at an appropriate level of significance with n-k (number of flows-number of coefficients) degrees of freedom, the model can be rejected as not providing an acceptable fit of the data. Alternative models can therefore be compared by assessing their deviance values. If an additional variable is introduced and there is an overall reduction in deviance, the latter reduction gives a measure of the importance of the new variable. It is usual in this type of modelling to begin from the null model (or grand mean model) in which M_{ijast} is estimated as simply the average flow (or rate) in the system. This serves as the baseline against which the results of other models (incorporating coefficients to identify effects of origin, destination, age, sex and time) can be compared. The deviance is also equal to the entropy statistic, which is used frequently in spatial interaction modelling.

Van Imhoff *et al.* (1997) calibrated models for the Netherlands, Italy and the UK specifically for the purpose of investigating to what extent the full multi-dimensional migration matrix could be simplified without seriously affecting the performance of the model. Their results indicate that a model of reasonable fit should contain at least the following interactions: origin-destination (OD), age-origin (AO), age-destination (AD) and sex-age (SA). In other words, the best model requires interactions among age, sex and origin and similarly between age, sex and destination, but the origin-destination effects are independent of age and sex. It was also found that time interacts with the main effects only (i.e. with age/sex, with origin, and with destination) and the remaining components (e.g. age/sex origin, age/sex destination, origin/destination, can be held constant). The absence or presence of interactions with the time dimension is crucial for using the model in internal migration projections. For making assumptions about internal migration the time invariant components need not be taken into account, and explicit hypotheses are only necessary about the time varying components. In a subsequent article (Van der Gaag *et al.,* 2000) explicit hypotheses were made about each of the time interactions in the model for projection purposes. For the time trend of the origin effects O as well as for the time trend of the destination effect D, three scenarios were proposed: (1) convergence, (2) divergence or (3) status quo. Convergence in the origin dimension implies that all origin-specific out-migration rates converge towards a common level, whereas divergence implies the reverse process in which existing differences become larger. Convergence in the destination dimension implies that the attractiveness of all zones converge towards a level which is proportional to their population size; divergence implies the opposite process whereby existing differences, standardised by their population size, enlarge. These convergence-divergence scenarios were used in the sub-national population projections for the European Union at the NUTS 2 level from a 1995 base.

2.6 Conclusions

The following conclusions have been drawn from this review of migration modelling.

Macro not micro: The review indicates that there is a long tradition of modelling internal migration and a wide variety of approaches that can be differentiated into those based on micro or individual decision-making and those that deal with macro effects on aggregate flows. It is macro approaches to migration modelling that are applicable in the context of this study.

Two-part modelling: State-of-the-art models divide the migration process into two parts: the first part models the migration out of each origin region; and the second part models destination choice. Certain models generate a pool of out-migrants which are allocated to different destinations, but it is preferable to allocate migrants from each origin to each destination because there are frequently important factors that link certain origin and destination pairs.

Determinants and selective influences: It is important to recognise the difference between those causal factors or variables that determine migration (such as marriage or job opportunities) and those factors that have a selective influence on migration (such as age or social class). It is essential to develop models of out-migration and destination choice that are age-specific and which divide the aggregate flow into appropriate life course groups. Sex is less important but should be incorporated if possible but there are likely to be severe data constraints on any further disaggregation by composition.

Diversity of explanatory variables: Migration flows, even of specific age groups, involve bundles of individuals motivated to migrate between regions for different combinations of reasons. Gravity variables that include the size of the origins and destinations and the intervening distances between origins and destination have proven to be important determinants in past studies but statistical relationships (signs, significance) between migration and many explanatory variables (unemployment, wages) have turned out to be specific to the system of spatial units being used and the national socio-economic conditions prevalent at the time. It is tempting to try and build a model containing a large number of explanatory variables but this makes huge demands on data collection, problems of autocorrelation and lack of clarity in interpretation.

Zone size characteristics: It is a well-known axiom of migration that more people migrate over shorter distances than longer distances. Consequently, zone size is very important since systems with smaller zones are likely to pick up more residential migrants who are not changing their jobs and will have very different motivations from those moving job as well as house. The NUTS 2 regions across EU countries do differ significantly in terms of area and population size and this has implications for the variables that should be included in a general model; i.e. it is very unlikely that the same model variables would be applicable across all member countries.

Demographic approaches: From the developments in multi-state demography has emerged another genre of modelling internal migration that seeks to identify those demographic influences which have an important influence on the stability of migration flows over time and which distinguishes those direct and interaction effects between origin, destination, age, sex and time dimensions that are most important and should be incorporated within a general model, even though different variants are applicable in different countries.

Model formulation: Approaches to macro migration modelling have various alternative formulations and make use of different mathematical or statistical calibration techniques. Several studies have emphasised the benefits of the use of the general linear modelling approach in fitting explanatory models of migration and the application of the Poisson model has been used both in modelling sub-national migration to explain the relative effects of exogenous explanatory variable on out- or in-migration, and on origin-destination migration. It has also been used to investigate the importance of using fewer parameters than those suggested in a full multi-state model and of using migration flows dependent on other endogenous variables, for instance population size or composition. Standard software (GLIM) is available to calibrate Poisson models although there may be some difficulties in using this package when dimensions are big and the number of cases are being modelled is enormous.

Need for integration: It is clear that internal migration is influenced by various explanatory determinants and that demographic dimensions such as age and sex are important selective influences, but it is also clear that migration is a phenomenon that experiences historical dependence. The two types of migration modelling that have been identified in this review, gravity-based models and demographic models, might be usefully brought together to provide a hybrid approach that allows the impacts of both dimensions to be evaluated.

Explanation or projection: Modelling historical flow patterns and projecting what will happen in the future should not necessarily be considered to require the same model. A good explanatory model of migration distribution probabilities may prove much less effective in a projection context, in comparison with a model based on historical flows, for example, simply because of the inadequacies of the projection of the independent variables. However, one of the key features of a projection model may be to test out how sensitive migration is to policy measures such as job creation or house-building programmes. Consequently, experience suggests that a modelling system would be particularly useful if it provided users with the means to experiment with alternative scenarios based on policy related variables whilst also allowing for results to be simulated under a 'do nothing' assumption. Furthermore, some testing of a projection model against observed data should be undertaken where data permits.

3 Current practice in the European Union

For the development of a new European-wide sub-national migration model the latest information and experience with models of sub-national population projections and internal migration used in the countries themselves is a necessary prerequisite. The main aim of this chapter is to present how internal migration has been treated in sub-national population forecasts in the current Member States of the European Union (EU), and to draw some lessons for population forecasts for the EU from the models designed by the national statistical offices (NSOs). In order to assemble the information required we asked national forecasters in 12 countries of the EU (all except Denmark, Ireland and Luxembourg) to fill in a questionnaire describing how sub-national projections are prepared in their countries. This questionnaire updates two previous inventories: the Eurostat/NIDI inventory of regional projection model practice in EU countries (van Imhoff *et al.,* 1994; Van der Gaag *et. al*., 1997b), and the Council of Europe/Leeds inventory of internal migration information in 18 European countries (Rees and Kupiszewski, 1999a; 1999b). Since no NUTS 2 classification exists for Luxembourg and Denmark, these countries are not included in the study. Although for Ireland a NUTS 2 level comprising two regions has been introduced in 1998, this country is also excluded as no long-term time series at NUTS 2 level are available. For the United Kingdom, NUTS 2 compatible population projections are produced by the individual home countries; therefore England, Scotland and Wales are treated as separate entities. In addition to Belgium, Flanders (the Dutch speaking part of Belgium) has been added, as Flanders sets up its own projection. Whilst we have obtained information from all countries, France did not reply to the questionnaire but provided some general information on the model used, and Greece reported that they do not prepare projections at a sub-national level. Consequently, France and Greece are not included in the remainder of this chapter.

In the following sections the results of the questionnaire are described that relate to the methodology of sub-national migration projections and the way internal migration is treated within these projections. For more detailed information, we refer to the working paper by Kupiszewski and Kupiszewska (2003) in which the full results of the questionnaire are presented.

3.1 General characteristics of sub-national population projections

In the current section, some general aspects of sub-national population projections are presented, which have been assembled in Table 3.1. This table contains, among other information, details of the nature of the calculations, the number of variants, the year of most recent projections, the regional classification and the general structure of the model.

The majority of the projections are labelled by NSOs as forecasts. Only the projection for Flanders is supposed to be a more analytical, what-if type of exercise. Spanish projections in fact ignore internal migration, and, as it was said in the questionnaire, may be deemed a forecast only in the absence of any significant migration. English, Scottish and Welsh projections are named as "trend-based projections, made on the assumption that no significant change in trends occurs".
The majority of the countries do not calculate any variants. Finland compiles projections with and without migration. In Italy, the Netherlands and Sweden, a main scenario and two variants are considered. In Austria, several combinations of medium, low and high variants of fertility and migration have been used, resulting in 13 different variants.

Most projections are fairly recent, based on benchmark population in 2000 or later. Germany, Wales and Portugal use slightly older benchmark populations (1999, 1998 and 1995 respectively). The period for which projections are made varies substantially, from 10 years for Flanders to 50 years for Austria, Belgium and Italy. Except for Italy and Portugal, all countries use single year projection intervals. Only Austria produces sub-national projections every year. The Netherlands and Scotland produce them every second year and the rest of the countries in intervals from 3 to 5 years, in some cases when the need arises.

Geographies used for sub-national projections vary substantially. Some countries use very small units, such as 1000 sub-municipal levels in the Netherlands or municipalities in Finland (448) and Flanders (308). Austria, Spain, Portugal and Italy, on the other hand, use the NUTS 2 regional division (9, 18, 7 and 20 regions, respectively). As expected, the geographies used are mostly determined by local requirements (labour markets, migration zones).

The methodology used in most countries is based on the cohort-component method. Four countries declare the use of a cohort-component model and another three use a multiregional cohort component model. Certain other countries created more complex methodologies, sometimes incorporating additional variables. Probably the most complex is the Dutch model, which uses a multiregional cohort component "engine", but controls its parameters with a number of external variables, such as: the labour market, school-supply, the housing market and distance. Most regional projections are consistent with the national projection; only for Flanders and Germany is this not the case.

From this short and simple description it can be concluded that there is a substantial diversification in the ways sub-national population projections are made.

Table 3.1: General information on the most recent sub-national population projections

Country	Nature of calculations	Number of variants	Frequency of updating the official regional population projections	Year of the most recent regional population projections	Period covered	Projection interval	Regional classification	Number of spatial units	Are the most recent regional and the national population projections consistent?	General structure of the model
Austria	Forecast	13[1]	Every year	2001	2001-2050	Single year	Bundesländer	9	Yes, mix top down/bottom up approach	Multiregional cohort component model[2]
Belgium	Forecast	1	Every five years, but not regular	2001	2000-2050	Single year	NUTS 3	44	yes	Component method[3]
Belgium - Flanders	What-if analysis, simulation	1		2000	2000-2010	Single year	Municipalities	308	Not linked	Cohort component model
Finland	Forecast	2[4]	Every three years	2001	2001-2030	Single year	NUTS 5, municipalities	448	Yes, bottom-up consistency	
Germany	Forecast	1	Irregular, usually every three years	2002	1999-2020	Single year	NUTS 3	440	no	Multiregional cohort survival model[5]
Italy	Forecast	3[6]	Every five years	2002	2001-2051	5 years	NUTS 2	20	Yes, bottom-up consistency	Multiregional cohort component model
Netherlands	Forecast	3[7]	Every two years	2001	2001-2030	Single year	Submunicipal level, Municipal level; NUTS 3 level. Interregional migration modeled at NUTS 3 level.	Submuni-cipal level: 1000; Municipal level: 500; NUTS 3 level: 40	Yes, top down	A hybrid form of multistate cohort survival model[8]

Country	Nature of calculations	Number of variants	Frequency of updating the official regional population projections	Year of the most recent regional population projections	Period covered	Projection interval	Regional classification	Number of spatial units	Are the most recent regional and the national population projections consistent?	General structure of the model
							Intraregional migration at Submunicipal level.			
Portugal	Forecast	1	Every five years	1997	1995-2025	5 years	NUTS 2	7	Yes, bottom-up consistency	Sequential model[9]
Spain	Projections based on recent trends	1	Varied, depending on the difference between projected and observed figures on birth, death and migration. When figures from a new census are available projections are also revised	1995 revised in 2001	15 years	Single year	NUTS 2[10]	18	Yes, population of each region and national population are projected separately; afterwards, differences between total and regional projected populations are adjusted	Component model[11]
Sweden	Forecast	3[12]		2002	2001-2040	Single year	Local Labour Markets can be aggregated into NUTS 2	110	Yes, there is a yearly consistency adjustment to the national forecast	A pure demographic model[13]
United Kingdom - England	Trend-based projection, made on the assumption that no significant change in trends occurs	No, but in the past ad hoc scenarios have been processed	Long term projection every 3 to 5 years, Short-term sub-national projections are also produced about every two years when no long-term projections are	1998 long-term projection 2002 short term projection	1997 - 2021 for the long term projection, 2001 - 2010 for the short term	Single year	Local and health authorities in existence on 1 April each year[14]	380 local authorities	Yes, bottom-up constrained by national population projections	Cohort component model

Country	Nature of calculations	Number of variants	Frequency of updating the official regional population projections	Year of the most recent regional population projections	Period covered	Projection interval	Regional classification	Number of spatial units	Are the most recent regional and the national population projections consistent?	General structure of the model
			being produced and when national projections are available based on the same year		projection					
United Kingdom-Scotland	Projections based on recent past trends		Every two years	2000	2000-2016	Single year	Council Areas, Health Boards	32 Council Areas, 15 Health Boards	Yes, top-down	Component method
United Kingdom-Wales	Projections based on recent past trends		Every two years	1998	1998-2023	Single year	Four regional groupings of unitary authorities	22 local authorities aggregated to 4 regional groupings	Yes, bottom-up constrained by national population projection	Cohort component model

Source: Questionnaires filled in by representatives of the relevant national statistical offices.

[1] 9 variants: all combinations of 3 variants (medium, high, low) for fertility and migration, 1 variant with constant fertility (medium mortality, medium immigration), 2 additional mortality variants (high, low) with medium fertility and migration, 1 benchmark variant (all constant), internal migration rates are constant in all variants.

[2] LIPRO 4.0 model is used. Step 1: Calculates projection for Austria; Step 2: calculates projection for nine NUTS 2 regions (bottom-up); Step 3: Corrects sums of NUTS 2 for population by age and sex, births and deaths and migration flows with the projection for total of Austria.

[3] Component method : Calculation of the deaths; Calculation of the births; Calculation of the internal migrations; Calculation of the international migrations; Taking into account the nationality changes (naturalisations, regularisation's)

[4] Calculation with and without migration

[5] Exogenous rates of fertility, mortality, mobility on the bottom level. International migration is the only top-down element of the model. The exogenous total number of immigrants is distributed to regions and sex/age groups. The out-migration is calculated by rates of mobility of regional population

[6] Assumptions are adopted on mortality, fertility, out-migration, internal migrations. 1 main scenario and 2 variants (low and high) are considered. Main variant: O/D matrix of projection probabilities of migration is kept constant until 2050 at estimated values for 1997-1999. This means that total flows at interregional level may vary over time depending on the structure and total regional populations. High variant: the O/D matrix changes over time until 2010. Some O/D effects between regions increase by 5% in 10 years. Low variant: the O/D matrix changes over time until 2010. Some O/D effects between regions decrease by 5% in 10 years.

[7] Variant 1: Most likely trend; variants 2 and 3: High and low variants: uncertainty interval = trend + / - 2/3.

[8] A hybrid form of multistate cohort survival model, using non-demographic information to calculate values of parameters for the model. Non-demographic information used: labour market, school-supply, housing market, distances. The model performs the following steps: 1. Input of national population forecasts; 2. Calculation of region-specific immigration, fertility, mortality; 3. Application for each region (i.e. submunicipal district) of a dynamic household model, which gives starters, 'stoppers", those who want to

change residence because of household reasons; 4. Application of a schooling migration module, a work related migration module, another migration module that gives a pool of potential migrants by motive for each departure district; 5. Confrontation of each migration type with supply. First: calculation of interregional moves, in order to generate vacancies. Then: combining interregional and intraregional moves in the housing market demand-supply module:
immigration: distribution of housing demand over regions
starters per submunicipality;
house changers per submunicipality
schooling: supply = demand, no constraints
labour market: workplaces through exogenous economic module, those who leave the labour market, and job-changers; iterative approach of spatially matching demand and supply, taking into account unemployment. Gives interregional migration due to work reasons. These movers are added to the housing demand in the work region;, but may end up in adjacent regions because of housing market shortages; see below;
other migration: distribution over regions.
Demand and supply are matched iteratively in order to clear labour market and housing market. After 6 iterations calculation of unmet demand which is transferred to t+1
[9] The components of growth are projected one at a time, in a fixed sequence. Population projections for the NUTS 2 units, is constrained by the national population forecast(s) using the "bottom up" approach. The base year population was derived from the most recent census (1991). Mortality assumptions are based on the latest available life table, by age (five years age group) and sex, and by sub-national units (NUTS 2). Fertility assumptions are based on the analysis of past trends, in particular on the TFR from which age specific fertility rates are extrapolated. Immigration and emigration are both expressed in absolute numbers and distributed by sex and five-year age groups.
[10] For NUTS 3 estimated figures by sex and five year age groups are avaulable.
[11] Components method: $P_{(t+1)}= P_{(t)}+N_{(t)}-D_{(t)}+I_{(t)}-E_{(t)}$
Initial census population by sex and age at the end of year t.
Surviving population from initial population at the end of t+1 year .
Addition of international immigrant flows during year t and surviving population at the end of the year.
Addition of internal migration (arrivals and departures) by age and sex during year t.
Projected births for the year t and survivals at the end of the year.
[12] Mainly based on various assumptions (historical trends on internal migration)
[13] A pure demographic model on the bottom-up basis, yearly modified to be consistent with the official national population forecast carried out by Statistics Sweden.
Region-specific population by sex and one year age groups are used as a starting point. National forecasts on death and birth rates have been adjusted by regional variations in terms of regional indices. Region-specific out-migration rates and in-migration distribution based on various historical patterns are used.
[14] Published projections are reaggregated and published to take account of changes.

- Projections are produced for local and health authority areas of England:
- Government Office Regions;
- counties and unitary authorities;
- local authority districts and London Boroughs.

3.2 Internal migration data: definitions and availability

It is important to make a distinction between various types of migration data. Information typically derived from the records of population registers counts all changes of address, sometimes conditional on crossing an administrative boundary. That means that all *events* (migration) are recorded and one may experience more than one *event* in a given period of time. If a migrant makes several migrations over a period of measurement, each of them will count separately. The death of a migrant has no influence on the migration count if his migration(s) took place between the start of a period and the time of death. If a migrant was born in the middle of a period of measurement and subsequently migrated, his migration will count as well. Return migration will count as two independent migrations. Registration is the most exact form of gathering data on migration. Later on we will refer to this data as movement data.

The transition type of data on migration is obtained by comparing places of residence in two points in time. This information is often collected during censuses of population by asking a question on the place of residence either at a specific date e.g. at the time of previous census or some (often one or five) years ago. This allows for cross-tabulation of places of residence at the beginning and at the end of the period specified in the question. It captures the aggregated result of all migrations of an *individual* over a period of time, irrespective of the actual number of migration (*events*). It does not capture return migration at all (from i to j and then from j to i) if they occurred over the period covered in the question asked. Neither does it include the mobility of persons who had migrated and subsequently died during this period. The migration of children who had been born and migrated during the period of measurement are sometimes accounted for. During census tabulation, infant place of residence at the start of the period of measurement is frequently assumed to be the place of birth. Multiple migration is not accounted for and is only represented as a transition resulting from the sum of migrations an individual makes.

The difference between the two types of data have been known for a long time (Courgeau, 1973a; Rees, 1977; Rees and Willekens, 1986). Unlike the movement approach in which the migration events are counted, the transition approach counts migrants – persons who in a given period migrated and survived on a given territory till the end of the period. The longer the period of measurement, the larger the difference as more multiple and return migrations (events) are ignored in comparison to the movement approach. The relationship between one year and five year migration data has been discussed in general terms by Kitsul and Philipov (1981).

Understanding the differences between the two types of data is very important when making population forecasts as each represents different measures of migration intensity and, as a consequence, requires different formulations of population dynamics models (Woods and Rees 1986). In the review of the data available, we did not examine the availability of other types of migration data, such as results of cross-tabulating the place of enumeration with the place of birth or previous place of residence, because this does not allow for capturing the mobility over a specific period of time.

Movement data availability

The definitions used in EU countries have been tabulated in Table 3.2. They refer to the data immediately available rather than to data available in principle (Rees and Kupiszewski, 1999b). The majority of countries use registration (movement) data. Portugal and the UK use census data (transition approach). The UK (England) projection model uses data from the NHS Central Register to update census migration. In order for migration to be counted usually the migrant has to cross an administrative boundary of a municipality. In some countries, as for example in the Netherlands or Sweden, the requirement is much weaker: a change of address is enough to count the migration. Mostly, there is no requirement to stay in a destination for a specific period of time.

Full flow matrices are available in 8 countries listed in Table 3.3. However the availability of age and sex details does vary. Belgium does not collect any age details; Germany collects data in broad age groups and the rest of the countries (Austria, Finland, Italy, the Netherlands, Spain and Sweden) collect data in single year age groups. These countries collect data up to 95 or even 100 years of age. The length of time series available varies substantially from the period from 1952 to 1999 in the case of Belgium to the relatively short period from 1996 to 2002 in Austria, whose model previously relied on census data. The UK is able to estimate full flow matrices. Germany has data available since unification, i.e. from 1991 onwards.

Table 3.2: Migration data types and definitions

	Type of data	Is a crossing of an administrative boundary required for a migrant to be counted?	Is it necessary in your definition of migration for a migrant to reside at the destination for more than a certain length of time?
Austria	Movement, Transition	Yes, municipality	No
Belgium	Movement	No	No, declaration of intention
Finland	Movement	Yes, municipality	Declaration of permanent migration
Germany	Movement	Yes, municipality (NUTS-4)	No
Italy	Movement, Transition	Yes, municipality	No
Netherlands	Movement	No, change of address	No, registration is crucial
Portugal	Transition	Yes, municipality	
Spain	Movement	Yes, municipality	No
Sweden	Movement	No, change of address	
UK – England	Transition	Yes	No
UK – Wales	Transition	Yes	No
UK – Scotland	Transition	Yes	No

Source: Questionnaires filled in by representatives of the relevant national statistical offices.

Table 3.3: Availability of full flow matrix on NUTS-2 level - movement data

	For what time span are the data available	Are the data available by the sex of migrants	Are the data available by the age of migrants	What is the last (oldest) age group for which the data are available?
Austria	1996 - 2002	Yes	Yes, single year age groups	95+
Belgium	1964 - now	Yes	No	
Finland	1987 - 2002	Yes	Yes, single year age groups	All ages
Germany	1991 to 1999	No	Yes, age groups: 0-17, 18-24, 25-29, 30-49, 50-64, 65+	65+
Italy	1952 - 1999	Yes	Yes, single year age groups	No upper limit because data are individual
Netherlands	1970 - now	Yes	Yes, single year age groups	99+
Spain	1988 - 2002	Yes	Yes, single year age groups	100+
Sweden	1968 - now	Yes	Yes, single year age groups	100+

Source: Questionnaires filled in by representatives of the relevant national statistical offices.

Data on arrivals and departures as well as on net-migration and the total number of migrants may be aggregated from the full flow matrix. It is therefore justifiable to look at the additional information available for these aggregations that are not available for full flow matrices (Table 3.4). In fact, such an additional dimension is available only in Belgium, which collects, since 1989, data on the age structure of migrants in 5 year age groups, with the oldest age group of 100+.

Apparently, except for some countries which have chosen to base their population statistics on censuses, there is very good provision of detailed movement data with almost universal availability of origin-destination-age-sex (ODAS) matrices in the last decade.

Transition data availability

The countries which collect transition data can be divided into two categories. Category one contains countries that either do not maintain population registers or do not use these registers for statistical purposes: France, Portugal and the UK. In the latter category, however, data from health administration are available and supplement the census data. In the second category belong countries which collect both registration and census data: Austria and Italy.

Typically the so called 1-year or 5-year questions are asked in the census. These questions are usually formulated: What was your usual place of residence one year ago/five years ago? The 5-year question was asked in Austrian and Italian censuses; other countries use a 1-year question (Table 3.5). The Netherlands uses a 2-year question in its periodic housing survey.

Predominantly, transition data come from censuses. Only data from this source are analysed in this section (Table 3.5). Three countries have a full flow matrix: Austria, Italy and the UK. Data for departures and arrivals are available for the same countries. In theory, transition data may be also derived from Labour Force Surveys. Surprisingly, this is not a popular option. It is unclear whether it is due to lack of reporting from NSOs or due to lack of interest in this type of data.

Apparently transition data are much less abundant than movement data and, as in the case of Austria and Italy, are collected in parallel to movement data. It seems the main body of information on migration comes from registration, and censuses play only a minor role. They may be, however, very useful in obtaining information not routinely collected by registration.

3.3 The use of internal migration data in sub-national population projections

The characteristics of internal migration data used in sub-national population projections are shown in Table 3.6. The strategies adopted by forecasters to model migration flows show a marked transition towards an information-rich environment. In most cases, a full matrix of flows has been used, considered indispensable for classical multiregional forecasting models (Rogers, 1995). Some countries which use a pool model for migration (Sweden, Spain) rely on data on departures and arrivals. In the case of Scotland, no intra-Scottish migration data are used; all migration flows are external. The Flemish model uses net-migration only. Migration data are often used either in 5 year age groups or in some other age aggregations, and estimates are made for single year age groups to meet the requirements of the models, which in most cases operate on single years of age. The use of time series for the estimation of model parameters (from 3 to 9 years) is widespread. The nature and scope of data on internal migration available in Europe is discussed in detail in Rees and Kupiszewski (1999b).

All countries with the exception of Germany, Belgium and the Netherlands use purely demographic models. The Dutch model uses non-demographic variables to control internal (in-migration and out-migration) and international migration (immigration only). Three groups of variables are used: variables depicting changes in the housing market, variables related to the labour market and schooling variables. The former are represented by additions and removals of housing stock, vacancies of housing created through moving households, supply-demand, and prices of housing and housing types (rental, owner occupied; single unit-flat). The following labour market variables are used: vacancies and job-seekers per region (NUTS 3); labour force participation rates and projection-based exogenous economic scenarios; variables describing the trade off between commuting and migration (distance dependent) and variables describing the trade off for the unemployed between staying unemployed and local demand for labour. Variables characterising the educational system are: higher education schooling facilities by region, forecasts of inflow into higher education by age, sex and municipality and historical flow patterns. Distance between regions is used as a geographical interaction-decay variable.

The German model has what we would call a distinct geographic flavour, as it looks at the saturation of the process of urbanization in former East Germany and also considers the consequences of liberalization of the labour market in Germany for the citizens of accession countries.

Table 3.4: Availability of migration movement data on NUTS-2 level: arrivals and departures

	Arrivals	For what time span are the data available	Are the data available by the sex of migrants	Are the data available by the age of migrants	What is the oldest age group for which the data are available?	Departures	For what time span are the data available	Are the data available by the sex of migrants	Are the data available by the age of migrants	What is the oldest age group for which the data are available
Austria	Yes	1996 – 2002	Yes	Yes, single year age groups	95+	Yes	1996 - 2002	Yes	Yes, single year age groups	95+
Belgium	Yes	1948 – now	Yes	Yes, five year age groups since 1989	100+	Yes	1948 - now	Yes	Yes, five year age groups since 1989	100+
Finland	Yes	1987 – 2002	Yes	Yes, single year age groups	All ages	Yes	1987 - 2002	Yes	Yes, single year age groups	All ages
Germany	Yes	1991 – 1999	No	Yes, age groups: 0-17, 18-24, 25-29, 30-49, 50-64, 65+	65+	Yes	1991 to 1999	No	Yes, age groups: 0-17, 18-24, 25-29, 30-49, 50-64, 65+	65+
Italy	Yes	1952 – 1999	Yes	Yes, single year age groups	No upper limit because data are individual	Yes	1952 - 1999	Yes	Yes, single year age groups	No upper limit because data are individual
Netherlands	Yes	1970 – now	Yes	Yes, single year age groups	99+	Yes	1970 - now	Yes	Yes, single year age groups	99+
Spain	Yes	1988 – 2002	Yes	Yes, single year age groups	100+	Yes	1988 - 2002	Yes	Yes, single year age groups	100+
Sweden	Yes	1968 – now	Yes	Yes, single year age groups	100+	Yes	1968 - now	Yes	Yes, single year age groups	100+

Source: Questionnaires filled in by representatives of the relevant national statistical offices.

Table 3.5: Characteristics and availability of census based transition data on NUTS 2 level

	Type of data	Question asked	What is the date of the census/survey from which data are available?	Are the data available by the sex of migrants	Are the data available by the age of migrants	What is the last (oldest) age group for which the data are available?
Austria	Full flow matrix, arrival vector, departure vector	Where did you live 5 years ago?	1991	Yes	Yes, single year age groups	95+
Italy	Full flow matrix, arrival vector, departure vector	Where did you live 5 years ago?	1991	Yes	Yes, single year age groups	100+
Portugal	Arrival vector,	Where did you live on a specific date (approximately 1 year ago and 5 years ago)?	1991 and 2001			
UK - England	Not available	What was your usual address 1 year ago?	2001			
UK - Scotland	Full flow matrix, arrival vector, departure vector	What was your usual address 1 year ago?	2001	Yes	Yes	90+
UK - Wales	Not available	What was your usual address 1 year ago?	2001	Yes	Yes	90+

Source: Questionnaires filled in by representatives of the relevant national statistical offices.

3.4 Internal migration assumptions

There is a considerable variability in the internal migration scenarios (see Table 3.7). Austrian, Belgian, one of the variants of Finnish and Swedish projections and the main variant of the Italian projections assume no change in migration intensities, i.e. constant age-specific migration rates. Trend extrapolation is used in Flanders and in Germany till 2005. In other cases a variety of changes are simulated: in Flanders, net rates are linearly halved over the period 2000 – 2010; in Germany, a set of factors, such as the saturation of the suburbanization in East Germany and a balance in East to West Germany migration is taken into account; Italian forecasters assume ±5% variation in certain elements of the OD matrix.

Belgium and the Netherlands use non-demographic variables in setting up internal migration hypotheses. The Netherlands uses regional economic growth variables, regional labour market variables and housing market variables as well as regional housing policies. In all countries except Spain and England, a bottom-up approach has been adopted in setting up the internal migration hypothesis (migration assumptions are formulated for the smallest spatial units and aggregated for larger units and the entire country). In Belgium and Germany, spatial aggregations have been made. Apart from the age and sex variables, in Belgium separate assumptions have been adopted according to nationality and in the Netherlands for institutional migration of the elderly and students. In Belgium assumptions are made on the flows between three categories of regions: urban, periurban and others.

3.5 Conclusion

Looking at current practice in EU countries, we may summarize that the cohort component method, and in particular its more refined variant - multiregional or multistate models - dominate the field of sub-regional forecasting. The modelling and forecasting of components of change is much less uniform. This is particularly true for the modelling of migration. It is also very characteristic that the use of non-demographic variables, especially for the allocation of migrants is not widespread, except for the Dutch model, which is based on a complex and sophisticated system of interactions between migration and a variety of housing and social variables.

If we divide all countries along two dimensions: 1) the type of variables used (demographic versus demographic plus non-demographic variables), and 2) the character of scenarios (no assumptions and status quo versus explicit assumptions), we will arrive at a simple classification. The Netherlands would be the only country which both uses non-demographic variables and sets up non-trivial migration scenarios. Germany could be counted among the countries which use non-demographic variables and makes only simple trend-based migration assumptions. Belgium also uses non-demographic variables in setting internal migration asumptions and sets up a status quo scenario for flows between various categories of regions. However, the use of non-demographic variables is limited to the internal migration scenario part of the model. Two other classes in which only demographic variables are being used have, roughly speaking, similar number of entries for both the lack of migration scenarios (status quo assumption and no migration assumption) cell and the migration scenarios cell. It should be noted however, that even if a scenario is adopted, it is rather a simple one.

The most important conclusion drawn from this inventory is that forecasters use very simple techniques of setting internal migration hypotheses. These techniques, except in a small number of cases, may be defined as status quo projections, no migration projections and linear trend projections. This is both good news and bad news. The good news is that there is plenty of room for improvement of internal migration forecasting. The bad news, however, is that despite abundant literature on causes and factors of internal migration, there is relatively limited expertise accumulated in the European forecasting community, which means that little experience-based input to such forecasting can be gathered locally

Table 3.6: Characteristics of data on internal migration used in subnational population projections

	Type of data	Intensity measure	Age groups	Highest age group	Aggregations	Smoothing	Other classification of the data	Time series used (first year – last year)
Austria	Full flow matrix	Age specific rates	Single year age groups	95+			Sex	1996 - 2001
Belgium	Full flow matrix	Out-migration rate	5 year age groups	85+	Clusters of arrondissements	Yes	Sex, nationality	1989 - 1997
Belgium - Flanders	Net-migration (including international migration)	Average age and sex specific net-migration rates at the level of municipality	Single year age groups	100+		Age profile smoothed over ages using running mean		1994 - 1996
Finland	Out-migration rates, probabilities of migration between major areas, in-migration shares				158 areas defined based on out-migration susceptibility of the population aged 15-44 years			1996 - 2000
Germany	Full flow matrix, shares of persons leaving place of residence to a specific destination	Migration rates, migration probabilities	Six broad age groups (0-17, 18-24, 25-29, 30-49, 50-64, 65+), combined with the estimation of single years using information of the top level	65+	No	No		1996 - 1999
Italy	Full flow matrix	Matrix of migration probabilities between regions (weighted averages of levels observed in 1997-1999 by age and sex)	Single year age groups	90+	Yes[1]	Yes		1997-1999

	Type of data	Intensity measure	Age groups	Highest age group	Aggregations	Smoothing	Other classification of the data	Time series used (first year – last year)
Netherlands	Full flow matrix	Out-migration rates, and in-migration probabilities conditional on out-migration	Six broad age groups (0-15-15-24, 25-34, 35-49,50-64, 65+), although in principle SN has 1 year interval data, and 5 year interval can be obtained at the regional level	65+		After modelling migration, the groups are redistributed over 1 year age intervals	Sex	1994-2001
Spain	Out-migration, in-migration	In-migration flows. out-migration rates	Single year age groups	100+			Sex	1986 - 2001
Sweden	Out-migration, in-migration	Out-migration rates, in-migration distribution	Single year age groups	95+		None		Various
United Kingdom - England	Full flow matrix	Occurrence – exposure rates	Single year age groups	85+		Rogers - Castro model for out-migration age distribution, gross migraproduction rates are used to estimate the level of migration	Sex	Combined 1992 - 1996 NHSCR data and 1991 census data

	Type of data	Intensity measure	Age groups	Highest age group	Aggregations	Smoothing	Other classification of the data	Time series used (first year – last year)
United Kingdom - Scotland	Net-migration (including international migration)	Net-migration	Migration assumption uses totals only. Model disaggregates according to 3 year average migration age distribution based on information from NHSCR controlled to national migration assumptions	Totals only in assumptions dissagregated to single age groups 90+ for councils and Health Boards			Sex	(base year -2) to (base year)
United Kingdom - Wales	Full flow matrix	Occurrence – exposure rates	Single year age groups	85+		Rogers - Castro model for out-migration age distribution, gross migraproduction rates are used to estimate the level of migration	Sex	Combined 1992 - 1996 NHSCR data and 1991 census data

Source: Questionnaires filled in by representatives of the relevant national statistical offices.

[1] Similar patterns by age were established for certain regions characterised by similar behaviour in terms of internal migration.
The groups are:

- Piemonte, Valle d'Aosta, Lombardia;
- Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia;
- Liguria, Emilia Romagna, Toscana;
- Umbria, Marche, Abruzzo, Molise;
- Lazio;
- Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna

Table 3.7: Characteristics of the scenarios on internal migration

	Number of scenarios/ variants	Characteristics of scenarios	Assumptions	Are the assumptions bottom-up or top-down	For which migration parameters are hypotheses made	Are hypotheses for separate groups of internal migrants distinguished?	Age groups	Are specific hypotheses for different groups of regions made?
Austria	1	Status quo	Internal migration rates are constant	Bottom-up	Out-migration rates and in-migration totals	Sex, age	0 - 95+	No
Belgium	1	Status quo	Internal migration rates are constant	Bottom-up	Out-migration rates	Sex, age and nationality	0 - 105+	Rural-urban (urban, periurban, others)
Belgium - Flanders	1	Trend until 2010	Age profiles of net-migration rates are stable over time. It was further assumed that these net-migration rates will be halved (in a linear way) in the period 2000-2010.	Bottom-up	Net-migration rates			No
Finland	2	Calculations with and without migration	Status quo and no migration		Total net-migration, out-migration rates, in-migration shares			
Germany	1	Trend until target year 2005 based on non-demographic information	Institutions receiving ethnic Germans will end their activities to 2005. Suburbanization process in East Germany will normalise after a period of very high mobility. Net East West migration will move towards a balanced situation.	Bottom-up	Out-migration rates, in-migration totals, migrant distribution	Yes	0-17, 18-24, 25-29, 30-49, 50-64, 65+	Yes, 5 clusters, exclusively estimated for the purpose to make assumptions of mobility

	Number of scenarios/ variants	Characteristics of scenarios	Assumptions	Are the assumptions bottom-up or top-down	For which migration parameters are hypotheses made	Are hypotheses for separate groups of internal migrants distinguished?	Age groups	Are specific hypotheses for different groups of regions made?
Italy	3	Main, high and low variants	Main variant: OD matrix of probabilities of migration kept constant until 2050 at values estimated for 1997-1999. High variant: OD matrix of probabilities changes over time until 2010. Some O/D effects between regions increase by 5% in 10 years. Low variant: OD matrix of probabilities changes over time until 2010. Some O/D effects between regions decrease by 5% in 10 years.	Bottom-up	Probabilities of migration between regions	Age, sex	Single year up to 90+	
Netherlands	3	Main, high and low variants	Most likely trend, high and low variants = 1±2/3 of the most likely trend value. Mix of policy driven and non-demographic information: internal migration is a result of housing market policies: where to build new houses of what type? These housing market policies are specified at the housing market regions at the level of municipalities.	Bottom-up	Out-migration rates, in-migration shares and migrant distribution	Sex, age, students migration, institutional migration of elderly. Each migrant group: age/sex reacts differently to housing market incentives and constraints. Students are a separate group within the model, as well as institutional households. For each of these groups separate uncertainty intervals apply.		When testing the model and validating it, it turned out that certain interactions were severely underestimated. For these interactions, specific parameters were included, to fill the gap. In the model it is assumed that these parameters will converge to 0 in the future. In this sense, some region-pairs have additional hypotheses.

	Number of scenarios/ variants	Characteristics of scenarios	Assumptions	Are the assumptions bottom-up or top-down	For which migration parameters are hypotheses made	Are hypotheses for separate groups of internal migrants distinguished?	Age groups	Are specific hypotheses for different groups of regions made?
Portugal	1			Bottom-up	Total net-migration	No		No
Spain	1	Trend	Age-sex profies of migration flows are stable over time.	Hybrid top-down and bottom-up	Out-migration rates, migrant distribution	Sex, age	Single year up to 100	No
Sweden	3	Status-quo and scenarios based on various assumptions (historical trends on internal migration)		Bottom-up	Net-migration, out-migration rates, in-migration rates, migrant distribution	Sex	Single year age groups	No
United Kingdom - England	1	Status quo		Hybrid - bottom-up and top-down	Total net-migration, migrant distribution	No		No
United Kingdom - Scotland		No sub-national migration in the model						
United Kingdom - Wales	1	Status-quo		Hybrid bottom-up and top-down	Total net-migration, migration distribution	No		No

Source: Questionnaires filled in by representatives of the relevant national statistical offices.

4 Regions, data and hypotheses

In studying internal migration, the regional classification within countries is very important. In case of a limited number of large regions, lower migration levels are expected compared to a larger number of smaller regions. Although the NUTS classification (Nomenclature of Territorial Units for Statistics) (European Communities, 1999) aims for a single uniform breakdown of territorial units for the European Union, regions in different countries, as well as *within* each of the countries, can differ considerably. This chapter of the report outlines the regional systems at NUTS 2 level in the countries that have been chosen as case studies for the modelling work (Sweden, the Netherlands, Spain and the United Kingdom), indicates some of the characteristics of the dependent migration variables, and discusses the selection and preparation of the independent or explanatory variables used in the study. Trends in out-migration are graphed and some of the primary migration flows for a recent time period are mapped in order to illustrate the nature of the flows that are used in the subsequent modelling work. Finally, an attempt is made to articulate the basic hypotheses that underpin relationships between the dependent variables and each of the determinant variables.

4.1 Regional systems

The variety of regional classifications across Europe makes it very difficult to compare migration levels and patterns between countries. Each country has its unique set of sub-national areas. Regions can differ significantly both in terms of size and structural characteristics and those differences may have implications for the measurement of migration. Large regions, for example, may subsume within their boundaries as intra-regional migrants many of the flows that might be inter-regional if the regions were smaller. Thus, for countries with only a relatively small number of large regions, a lower rate of inter-regional migration may be expected in comparison with countries that have a large number of small regions.

The current project was commissioned at the scale of NUTS 2 regions, a EU classification of territorial units used by the European Commission for policy formulation and funding allocation. Whilst one might expect the NUTS 2 regions to show some measure of uniformity across the EU, in practice the variation in the numbers and size of NUTS 2 regions varies considerably from country to country. Table 4.1 illustrates the variation in numbers of regions and in population sizes between the case study countries: Sweden, the Netherlands, Spain and the UK.

Table 4.1: Characteristics of the case study NUTS 2 regions, 1998

Country	Population size (millions)	Number of NUTS 2 regions	Largest NUTS 2 region	Population size (millions)	Smallest NUTS 2 region	Population size (millions)
Sweden	8.8	8	Stockholm	1.8	Mellersta Norrland	0.39
Netherlands	15.6	12	Zuid-Holland	3.4	Flevoland	0.29
Spain	39.4	18	Andalucia	7.1	Ceuta y Melilla	0.13
UK	59.1	32	London	7.2	Cumbria	0.49

Source: Eurostat

These four countries have been selected because they represent the diversity shown in Table 4.1 but also because they contain representation from across the European space. Their NUTS 2 regions are illustrated in Figures 4.1-4.4.

It should be noted that in the UK, the complete set of NUTS 2 regions is not used. Thirty two NUTS 2 regions are defined with Scotland and Northern Ireland being treated as single units. Boundary changes have also been introduced in the UK during the 1990s and the spatial units used are those in place before 31 December 2000. The boundary changes affect London, Wales and the South West in particular and mean that some adjustment of data is required to ensure consistency over time during the 1990s.

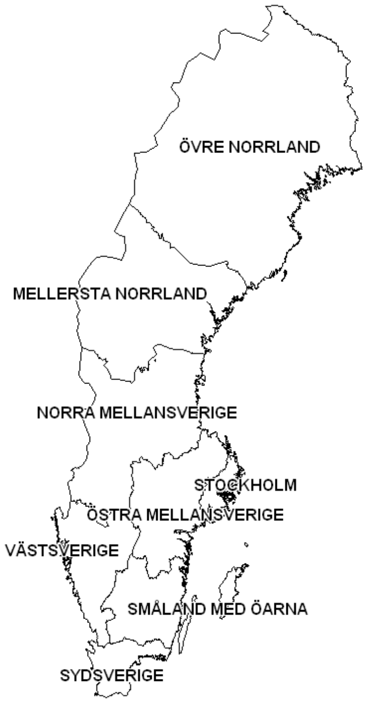
It is appropriate to recognise that the huge variation in population size that exists between and within countries is likely to have a significant impact on the modelling experiments and will make the search for a general model that much less straightforward. The selection of the four case study countries was partly determined by the availability of data and it is this issue to which we now turn our attention.

4.2 Internal migration data

Given that the project was specified at NUTS 2 level, it was initially planned to make use of data on migration flows from EUROSTAT. The following internal migration data are available in the REGIO database (*migr-r*):

- departures due to internal migration by sex and age (*p2dep*); and
- internal migration by sex, region of origin and region of destination (*p2mig*).

Figure 4.1: NUTS 2 regions in Sweden



1	se01	Stockholm
2	se02	Östra Mellansverige
3	se03	Sydsverige
4	se06	Norra Mellansverige
5	se07	Mellersta Norrland
6	se08	Övre Norrland
7	se09	Småland med öarna
8	se0A	Västsverige

Figure 4.2: NUTS 2 regions in the Netherlands

1	nl11	Groningen
2	nl12	Friesland
3	nl13	Drenthe
4	nl21	Overijssel
5	nl22	Gelderland
6	nl23	Flevoland
7	nl31	Utrecht
8	nl32	Noord-Holland
9	nl33	Zuid-Holland
10	nl34	Zeeland
11	nl41	Noord-Brabant
12	nl42	Limburg (NL)

Figure 4.3: NUTS 2 regions in Spain

1	es61	Andalucia
2	es24	Aragón
3	es12	Principado de Asturias
4	es53	Illes Balears
5	es7	Canarias (ES)
6	es13	Cantabria
7	es42	Castilla-la Mancha
8	es41	Castilla y León
9	es51	Cataluña
10	es52	Comunidad Valenciana
11	es43	Extremadura
12	es11	Galicia
13	es3	Comunidad de Madrid
14	es62	Murcia
15	es22	Comunidad Foral de Navarra
16	es21	Pais Vasco
17	es23	La Rioja
18	es63	Ceuta y Melilla (ES)

Figure 4.4: NUTS 2 regions in the UK

1	ukn	Northern Ireland
2	ukm	Scotland
3	ukc2	Northumberland, Tyne and Wear
4	ukc1	Tees Valley and Durham
5	ukd1	Cumbria
6	uke3	South Yorkshire
7	uke4	West Yorkshire
8	uke1	East Riding and North Lincolnshire
9	uke2	North Yorkshire
10	ukf1	Derbyshire and Nottinghamshire
11	ukf2	Leicestershire, Rutland and Northants
12	ukf3	Lincolnshire
13	ukh1	East Anglia
14	uki	London
15	ukh2	Bedfordshire, Hertfordshire
16	ukj1	Berkshire, Bucks and Oxfordshire
17	ukh3	Essex
18	ukj2	Surrey, East and West Sussex
19	ukj3	Hampshire and Isle of Wight
20	ukj4	Kent
21	ukk1	Gloucestershire, Wiltshire and North Somerset
22	ukk3	Cornwall and Isles of Scilly, Devon
23	ukk2	Dorset and Somerset
24	ukg3	West Midlands
25	ukg1	Herefordshire, Worcestershire and Warks
26	ukg2	Shropshire and Staffordshire
27	ukd3	Greater Manchester
28	ukd5	Merseyside
29	ukd2	Cheshire
30	ukd4	Lancashire
31	ukl1	East Wales
32	ukl2	West Wales and The Valleys

The departures data are available by five year age group but the data for 1995 are partial. Due to boundary changes there are several gaps in the data sets for Sweden and the UK at NUTS 2 level, and there are a lot of estimated values for all years for Sweden with no clear statement of the estimation methodology[1]. Origin-destination flow matrices are available at NUTS 2 level but these are incomplete for Sweden and non-existent for the UK (only at NUTS 1 level). More importantly, there are no age-specific data for any of the countries.

Given the drawbacks of the EUROSTAT data in REGIO, particularly the absence of age-specific data on origin-destination flows, it was decided that data sets should be obtained from the NSOs: Statistics Sweden, Statistics Netherlands, Insituto Nacional de Estadistica (INE) in Spain and the Office of National Statistics (ONS) in the UK.

Figures 4.5-4.8 contain sets of total out-migration flows and rates for the regions in each of the four countries respectively. All the data come from registration sources (not censuses or surveys) and they are included here to indicate the magnitude and temporal stability of total out-migration taking place during the 1990s. In Sweden, out-migration was falling in the initial years of the decade but this decline was reversed in 1993, peaked in 1994, dropped again and increased from 1996. The fluctuations experienced in Sweden contrast with the relative stability of the regional schedules in the Netherlands.

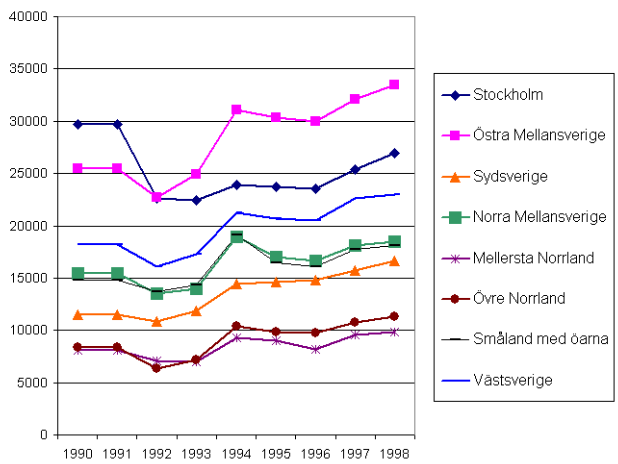
The upward trend in the published time series of Estadisticas Variciones Residentiales (EVR) data in Spain (obtained from INE) is interrupted in 1991 and 1996 because these are the years that coincide with a Census or Padron Municipal de Habitantes (PMH) which capture a considerable proportion of the migration registrations and reduce those enumerated in the EVR. Generally speaking, the levels of migration are much lower in Spain than elsewhere, particularly the UK, whose migration rates vary from below 0.5% per year to over 3.5% per year.

The UK data come originally from a central register of NHS patients re-registering with doctors in different regions and the data sets refer to 12 month periods from mid-year to mid-year rather than calendar year measures of migration as in the other countries. The data is made available as flows between Family Health Service Authorities (FHSAs) of which there are 98 in England and Wales. These flows have had to be aggregated to NUTS 2 regions.

[1] Although not of relevance for the current study, the outflows from NUTS 1 regions are aggregations of outflows from NUTS 2 regions. This might be confusing as out-migration is not a phenomenon that can be aggregated like population or employment because of the flows that take place between NUTS 2 regions within NUTS 1 areas.

Figure 4.5: Regional out-migration flows and rates in Sweden

Out-migration totals



Out-migration rates (%)

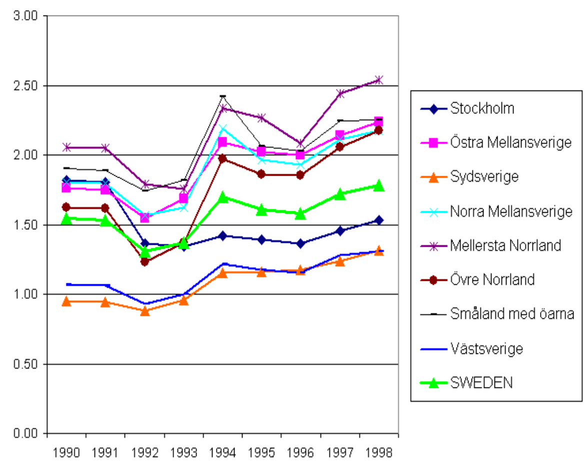
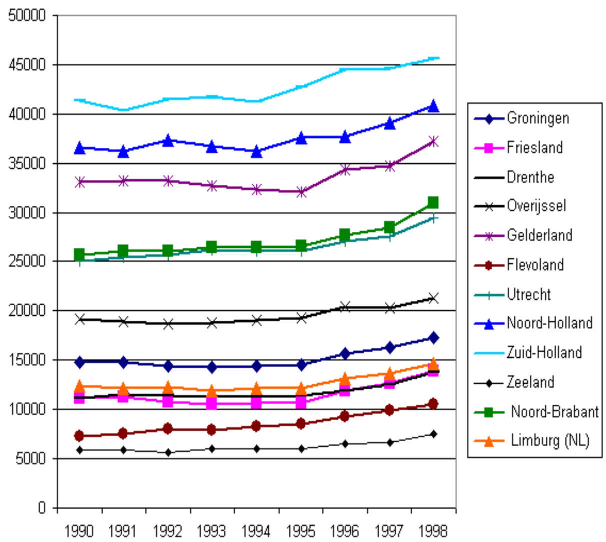


Figure 4.6: Regional out-migration flows and rates in the Netherlands

Out-migration totals



Out-migration rates (%)

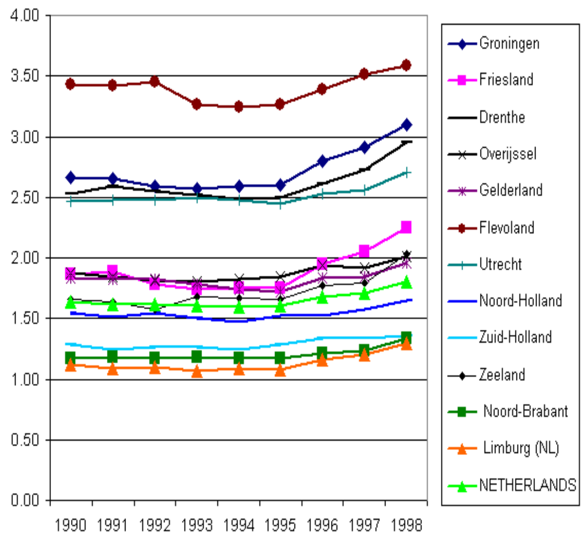
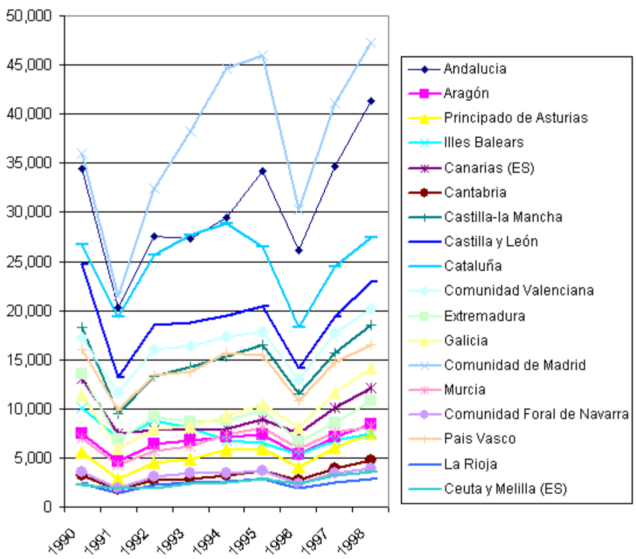


Figure 4.7: Regional out-migration flows and rates in Spain

Out-migraition totals



Out-migration rates (%)

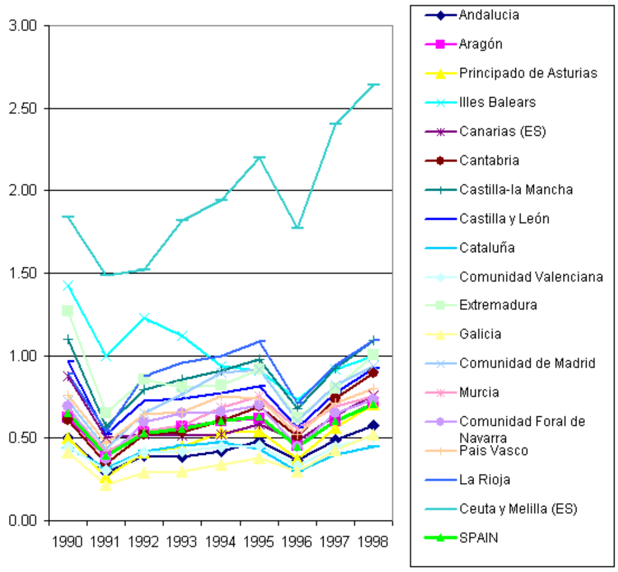
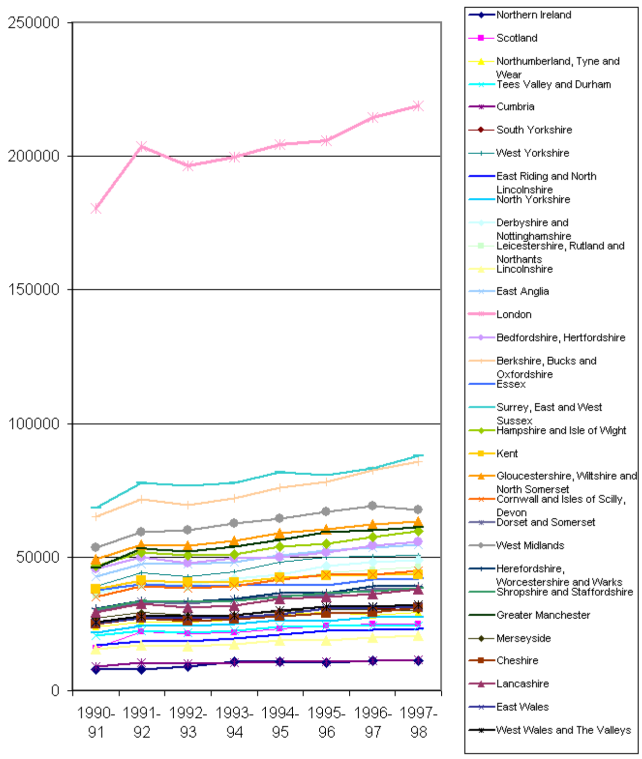
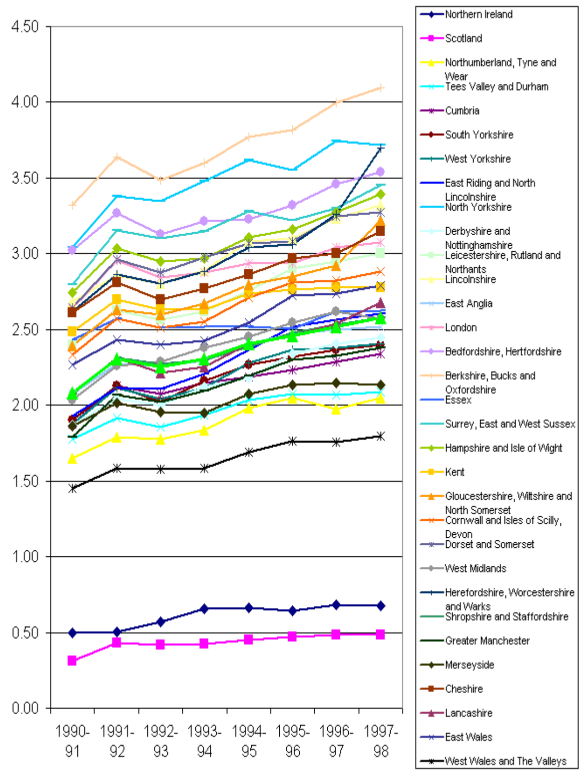


Figure 4.8: Regional out-migration flows and rates in the UK

Out-migraition totals

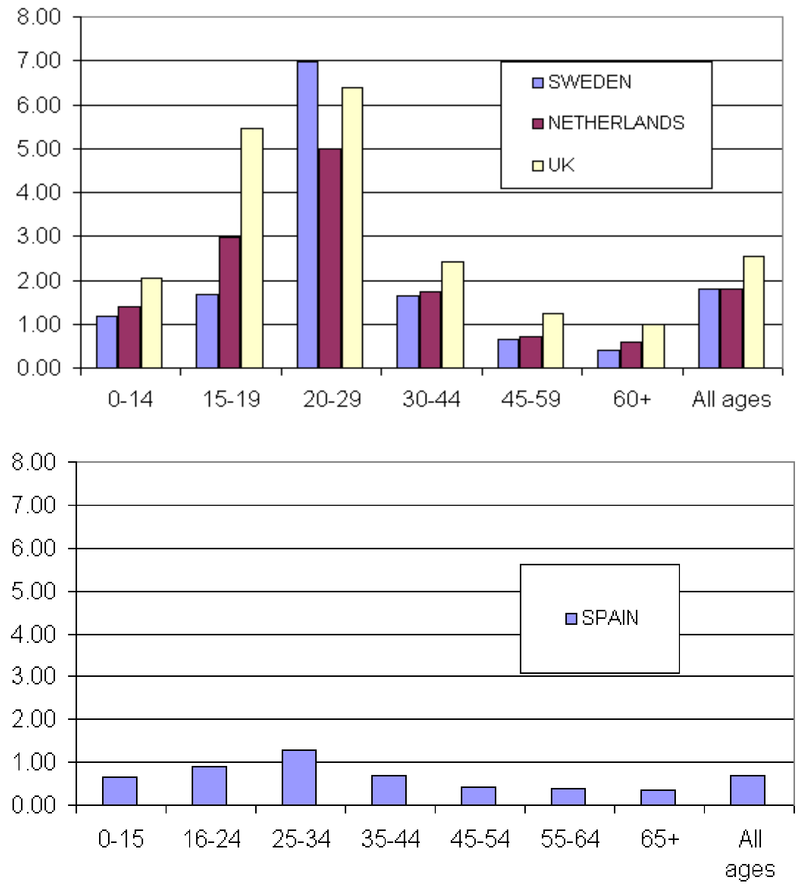


Out-migration rates (%)



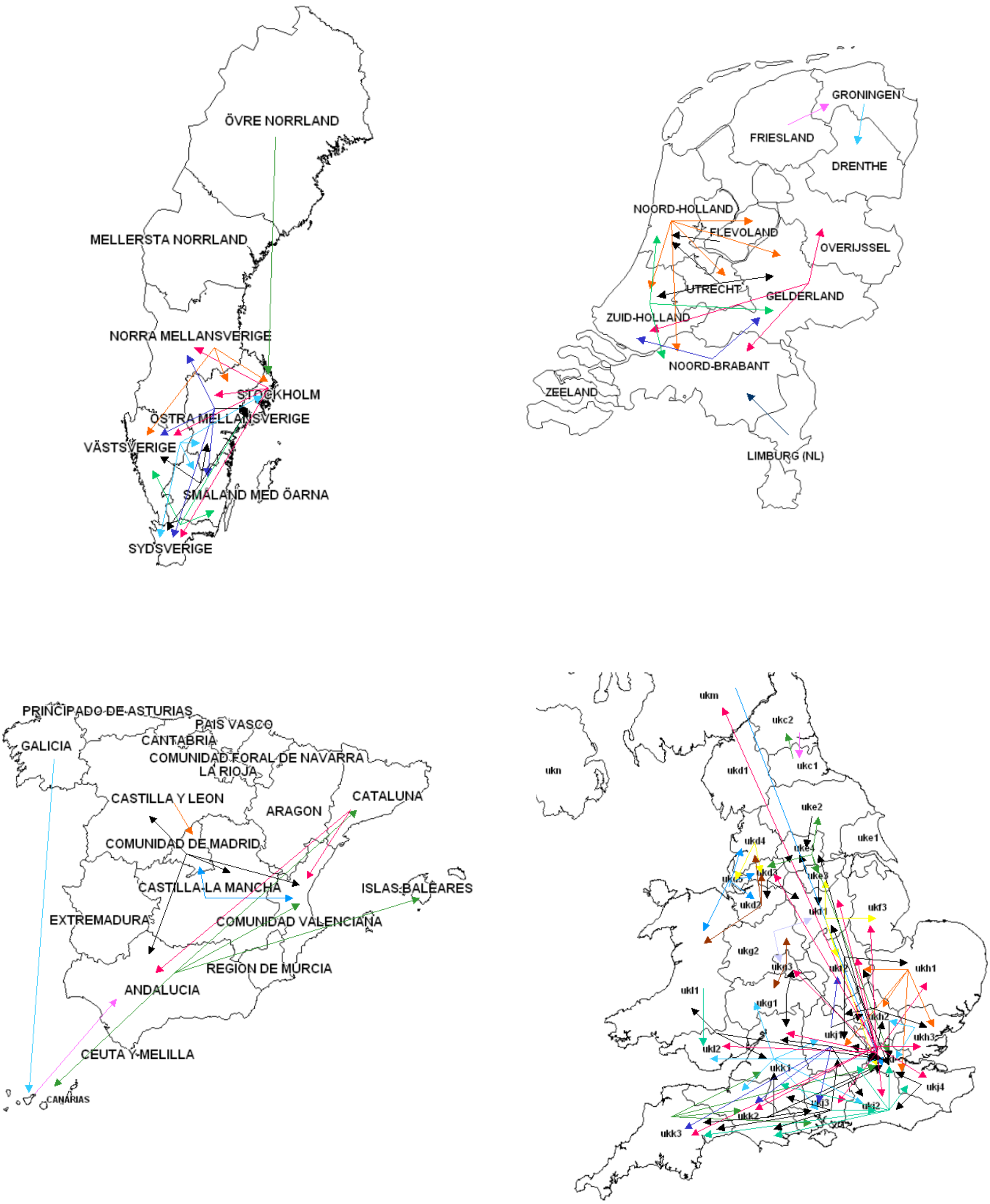
Inevitably, time series of age-specific flows and rates are likely to show greater fluctuation and the histograms in Figure 4.9 illustrate age rate differentials for migration in each of the four countries in 1998. A comparatively high rate of migration for those aged 20-29 in Sweden is apparent whilst the UK has the highest rates in the other age groups. Spain is plotted separately because the age groups are defined differently due to the lack of comparable population data to compute the rates.

Figure 4.9: Age schedules of out-migration rates (%) for 1998



In order to give an idea of the spatial patterns of migration between origins and destinations in each country, primary flows (those that involve over 3,000 persons) have been mapped in Figure 4.10. These maps give some indication of where the largest exchanges are taking place in 1998. In Sweden most large flows occur between regions in the south of the country and the largest flow of 3,500 occurs between Stockholm and Ostra Mellansverige. Primary flows in the Netherlands take place between the central regions with the largest (10,300) between Zuid-Holland and Noord-Holland.

Figure 4.10: Main origin-destination flows (> 3000 persons) in each of the case study countries in 1998



In Spain, the biggest flow (11,800) is between Madrid and Castilla La Mancha but there is more evidence of longer distance movements occurring to the islands and to and from Andalucia. Spanish data on age-specific origin-destination flows are available from the EVR records but only data for 1994 and 1998 was available for the project. The greater density of inter-regional migrations involving more than 3,000 people is evident in the UK where the largest flow is that between Greater London and Essex and involved 23,000 people in 1998. The importance of Greater London as both a generator and a recipient of migration is also demonstrated in Figure 4.10.

In summary, the migration data has been assembled into two files for each country (Table 4.2).

Table 4.2: Dimensions of the data files constructed

Case study	File	Origins	Destinations	Age	Sex	Periods	Records
Sweden	Out-migration	8	na	6	2	9	864
	Origin-destination	8	8	6	2	9	6,912
Netherlands	Out-migration	12	na	6	2	9	1,296
	Origin-destination	12	12	6	2	9	15,552
Spain	Out-migration	18	na	7	nu	2	252
	Origin-destination	18	18	7	nu	2	4,536
UK	Out-migration	32	na	6	nu	8	1,536
	Origin-destination	32	32	6	nu	8	49,152

na = not applicable; nu = not used

The final column of Table 4.2 shows the variation in the number of records in each of the files. The full set of dimensions (origin, destination, age, sex and time) is used in the cases of Sweden and the Netherlands. The data for Spain are only available for two time periods and sex has not been included. The size of the regional dimensions for the UK means that the origin-destination file contains nearly 50,000 records. The sex variable was excluded so as not to double the file size. The strategy for modelling set out later in the report does not involve calibration using all the records in each case. The files are split into two time periods in order to test alternative models based on the early 1990s against observed data for the late 1990s.

4.3 Explanatory variables

In Section 2.3 of this report, attention was drawn to the large number of causal factors reported in the literature and reviewed most recently by Champion *et al.* (1998). It was deemed necessary, therefore, to make some choices of explanatory variables and in this instance, pragmatism was a basic principle. It was assumed that:

- data should be available and consistent across countries wherever possible;
- a relatively small number of variables should be taken into account;
- both demographic and non-demographic variables should be used;
- no age or sex disaggregation of variables should be used;
- change variables as well as stock variables are important; and
- variables should be selected that can be used in a projection context.

Given the importance of consistency, it was decided to use data from EUROSTAT's REGIO database wherever possible, supplemented where necessary by data from national sources and/or by estimation. The demographic, economic and other variables selected are shown in Table 4.3. Information on a wider range of REGIO variables including educational performance, transport infrastructure and employment in high technology industry was considered but little information was available from the database for NUTS 2 regions.

Table 4.3: The variables selected for modelling

Demographic	Economic	Other
Population	GDP at ppp per inhabitant**	Distance
Density	Unemployment rate**	Contiguity
Immigration	Employment*	
Accessibility	Housing stock	

* Change variable computed
** Lagged and change variables computed

The REGIO data set for populations is *p2age90* and contains counts at 1 January each year. There are complete data sets for Netherlands and Spain but values are missing for the Swedish regions of Småland med öarna and Västsverige for 1990-1993 and for certain regions of the UK. Figures for Sweden were completed by using data of Statistics Sweden. For the UK, figures for London, Cornwall and Devon and two regions in Wales are taken from ONS mid-year estimates. An argument can be made for including a measure of demographic structure (e.g. dependency ratio) as an additional variable in the model but such a variable was not used.

Density values are available from *d3densit* in REGIO. These are computed as populations per square kilometre. Missing values for the UK are estimated by applying the rate of change between the two closest years in the time series.

Immigration data in REGIO is held in the *p2img* data set under *mig-r* Migration Statistics. Complete data sets exist for the Netherlands and Spain but some estimation is required for two Swedish regions and there are no data for the UK available here. The UK data come from various sources: for regions in England and Wales, ONS estimates from the International Passenger Survey (IPS) were obtained from the MIGMOD database; for Northern Ireland, estimates of flows from Republic of Ireland and from abroad were supplied by NISRA; for Scotland, estimates were obtained from the IPS, supplied by ONS.

Accessibility variables are not available from REGIO and are calculated using the population counts and the distances between the centres of each region in each country. The accessibility variable, defined as a measure of demographic potential, is calculated for each region i as $\sum_{j \neq i} P_j/d_{ij}$ and represents the nearness or proximity of the region to the existing distribution of population in the remainder of the country.

Data on GDP at purchasing power parities per inhabitant comes from the REGIO data sets *e2gdp79* and *e2gdp95* to cover 1990-98. Some estimation is required for the UK and Sweden: GDP values for London and the two Welsh regions for 1991-93 are estimated by applying region/national GDP proportions in the previous year; GDP for all regions for 1990 and for Småland med öarna and Västsverige in 1990-93 are estimated in the same way. The time series of GDP per inhabitant for Swedish regions show stability in the first half of the decade at around 15,000 and growth thereafter with clear differentiation between Stockholm and the rest of the country. The regions in the other countries, on the other hand, show increasing growth throughout the 1990s (Figures 4.11-4.14).

The REGIO data set used for unemployment rates is *un3rt*. Once again, the Swedish and UK regions need some estimation work to fill gaps in the time series. Swedish data on unemployment come from the Labour Force Survey (Table Befolkningen 16-64 ar (AKU) efter region, kön och arbetskrafsstatus, Ar 1976-2002) from the Statistics Sweden web site. The Swedish unemployment rate is unemployed persons expressed as a percentage of the active population (employed plus unemployed, ages 16-64). In the UK, there are some data available for 1990, 1991 and 1996-98, but data are sparse for the intervening years, 1992-95. Rates for NUTS 1 regions were assembled and used to estimate NUTS 2 regions for these years. Complete time series were estimated for London, East Wales and West Wales and the Valleys based on claimant counts for the constituent areas. The unemployment time series vary between countries. In Sweden, unemployment rates increase between 1990 and 1993 and then remain more stable with a decline towards the end of the period. In the Netherlands, rates fall from 1990 to 1993, increase to the mid-90s and decline thereafter. In Spain, which has by far the highest rates of all the case study countries, the trend is an increase to 1994 and then decline. This pattern is also applicable in the UK but the turning point is a year earlier (Figures 4.15-4.18).

Some data on employment is available from the *lf2emp* data set in REGIO, which is sourced from the Community's Labour Force Survey. However, figures for numbers employed are only available for limited number of years in the second half of the 1990s for Sweden, the Netherlands and the UK. Thus, the data has to be obtained from national sources where available.

REGIO contains no data on housing stock but data on the changing number of dwellings for years 1995-98 for the Dutch regions was obtained from Statistics Netherlands and data on new constructions (apartments and single family dwellings) for all years was available from Statistics Sweden. No data on housing completions were obtained for Spain or the UK.

Figure 4.11: GDP per capita, 1990-1998, Sweden

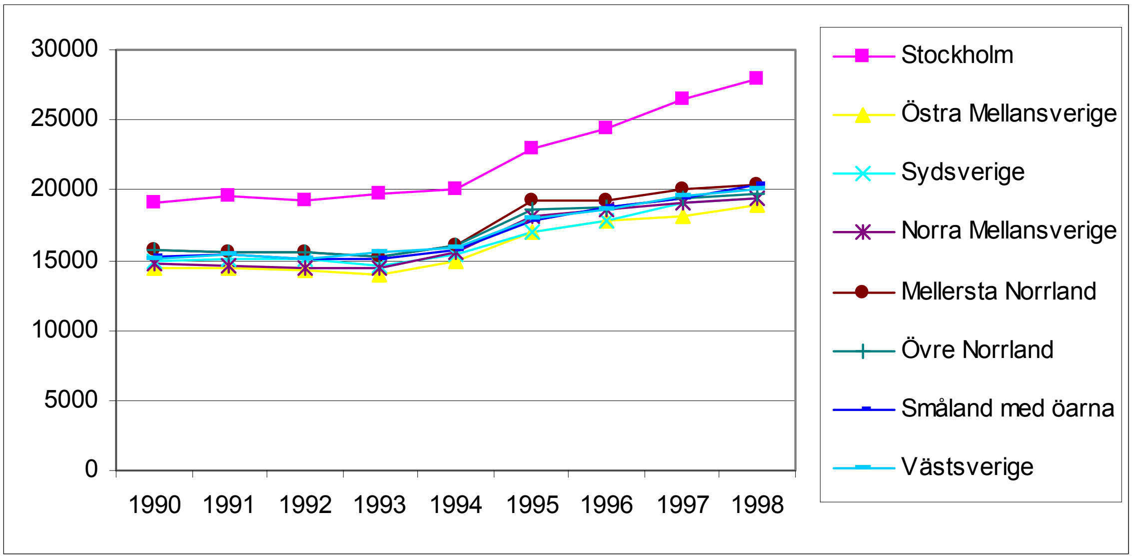


Figure 4.12: GDP per capita, 1990-1998, the Netherlands

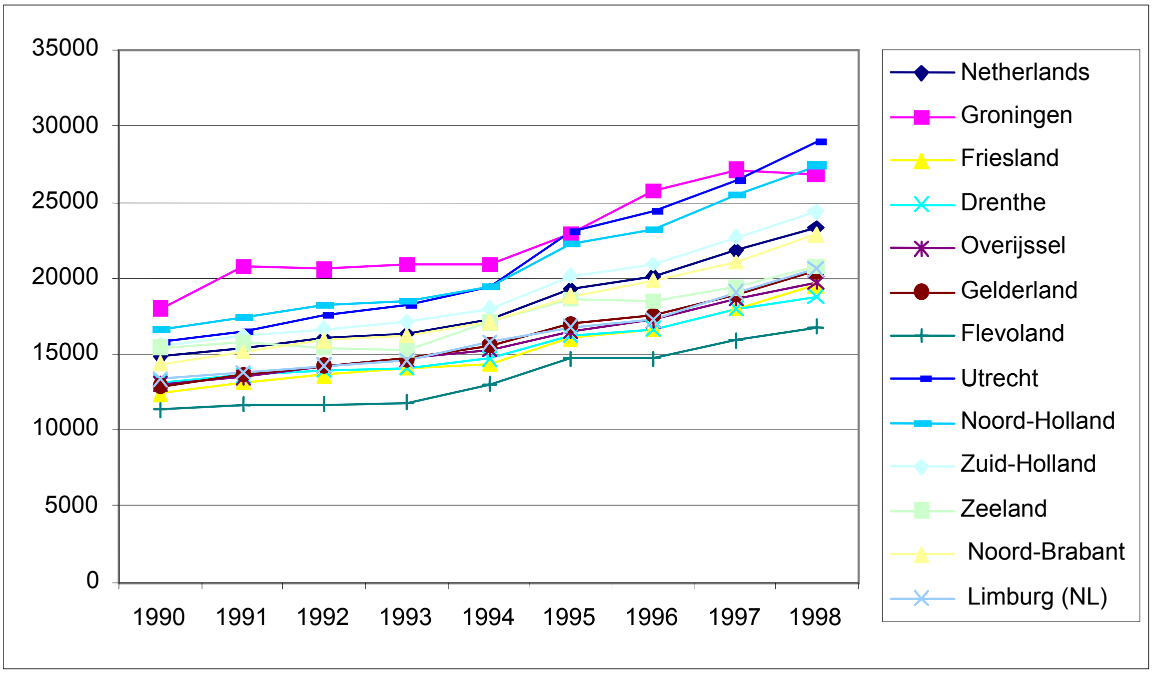
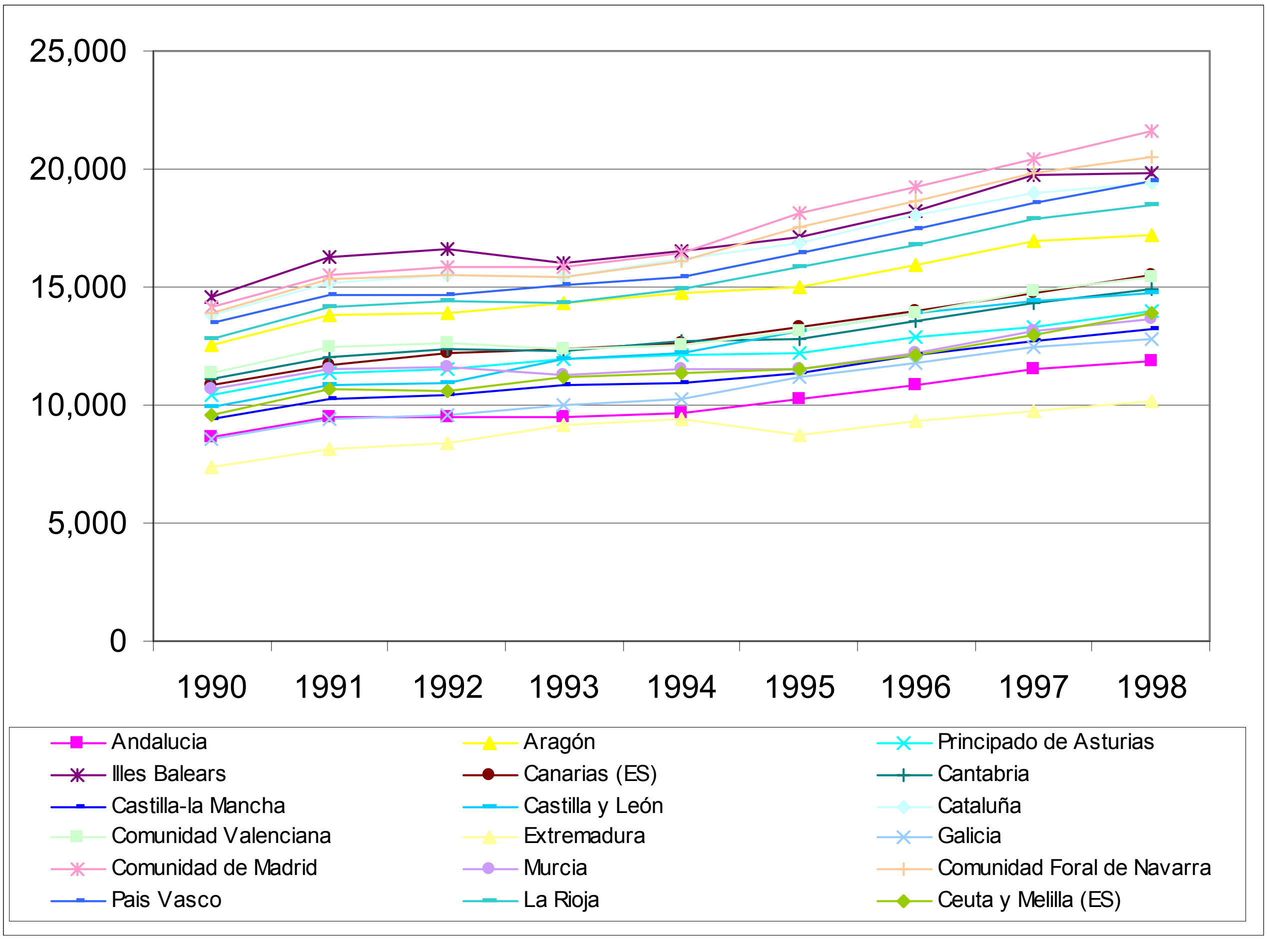


Figure 4.13: GDP per capita, 1990-1998, Spain



The distance variable was not available from REGIO or from national statistical offices. In the case of the UK, the matrix of distances (in kilometres) between NUTS 2 areas was built up as an average of distances between smaller spatial units used in the MIGMOD project. The compilation of the distance matrix for Spain was made more difficult because of the islands (Baleares, Canarias) and Ceuta y Melilla. Shipping distances were used for flows between these regions and to the capital cities of mainland regions. This methodology has been used previously by INE (INE, 1993, *Indicadores Sociales* 1991). For Sweden and the Netherlands, straight line distances were calculated between the centres of NUTS 2 regions.

Finally, the contiguity matrices of ones (indicating contiguity) and zeros (indicating no contiguity) were produced manually.

Figure 4.14: GDP per capita, 1990-1998, United Kingdom

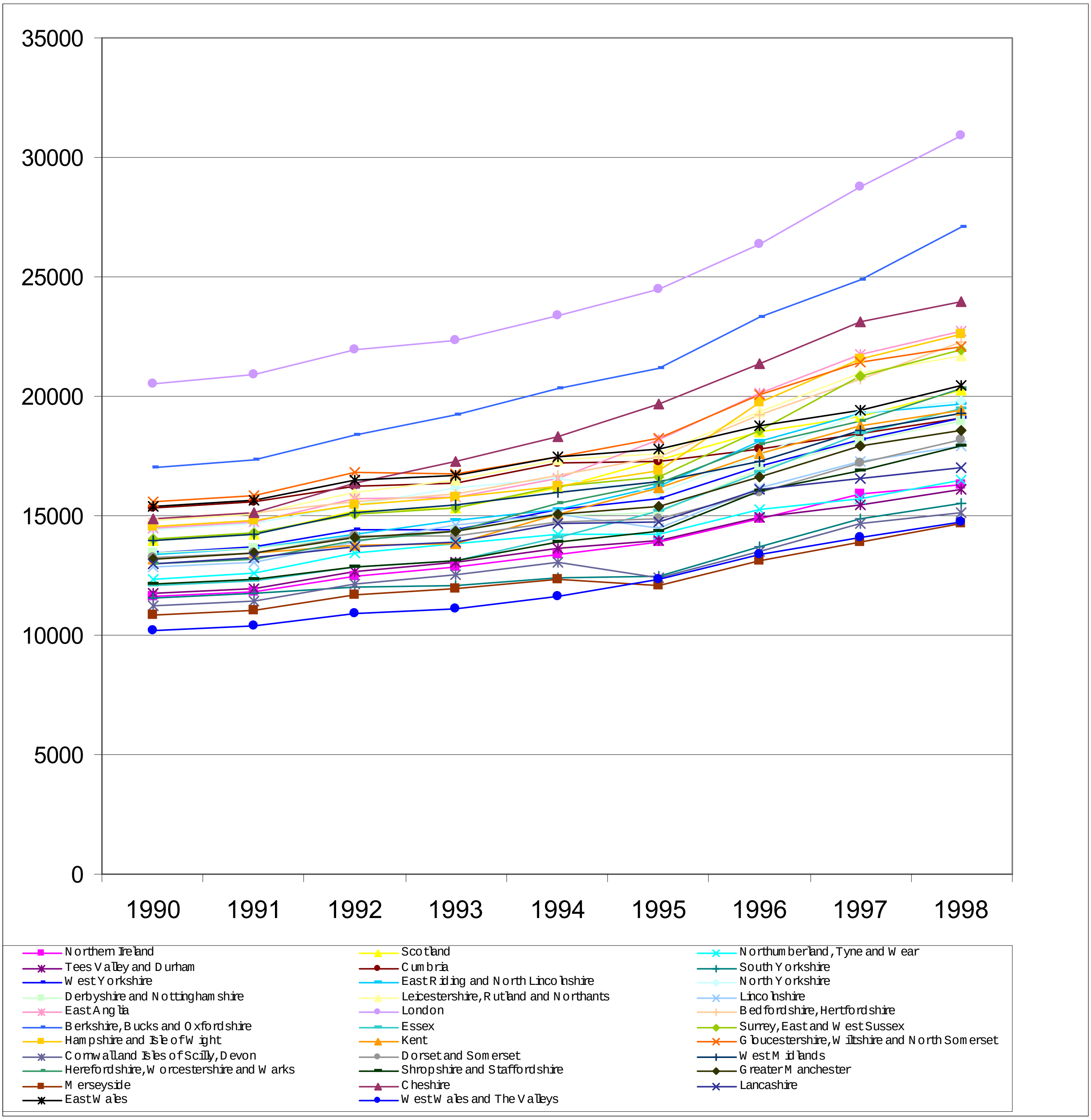


Figure 4.15: Unemployment rates, 1990-1998, Sweden

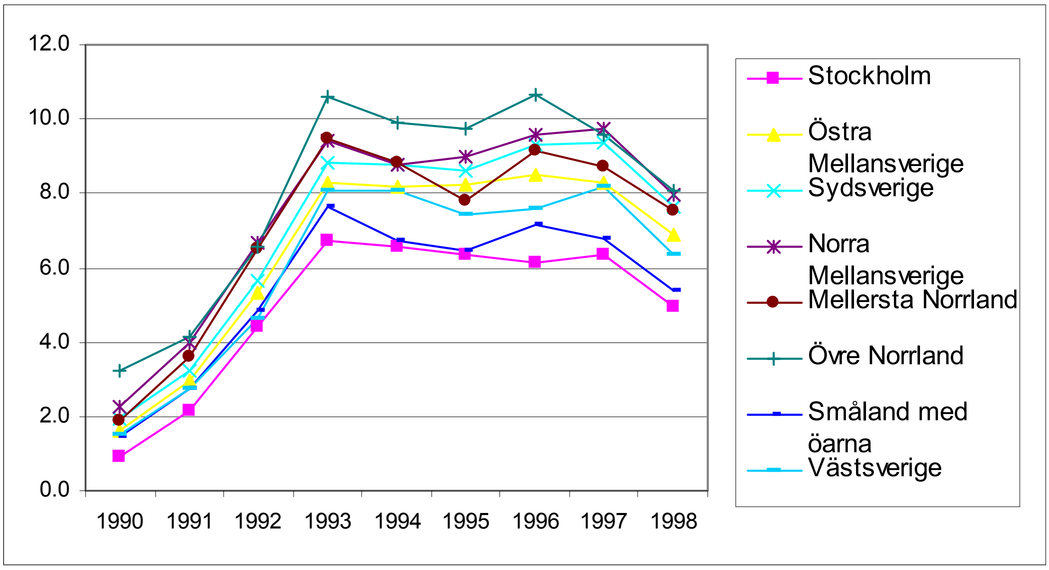


Figure 4.16: Unemployment rates, 1990-1998, the Netherlands

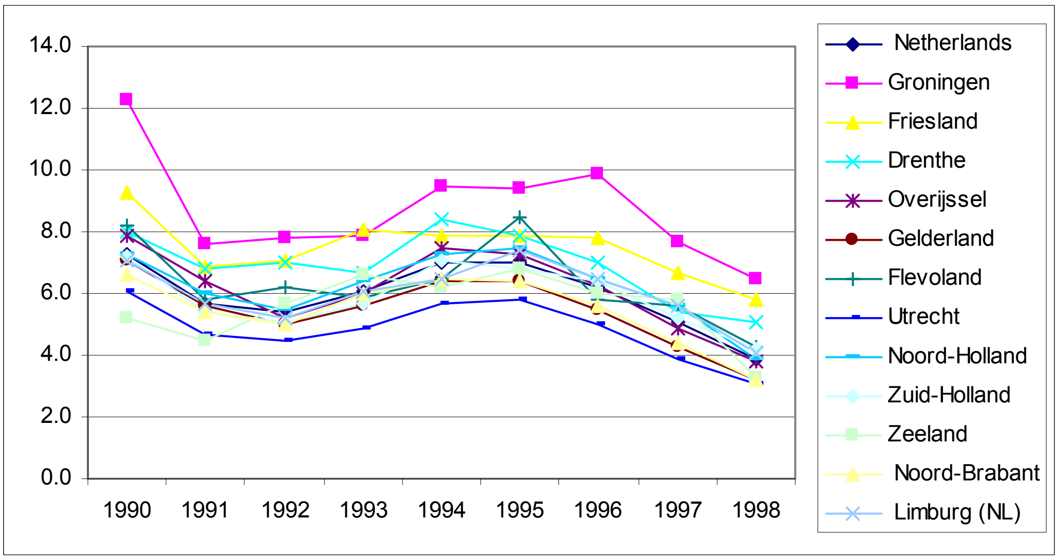
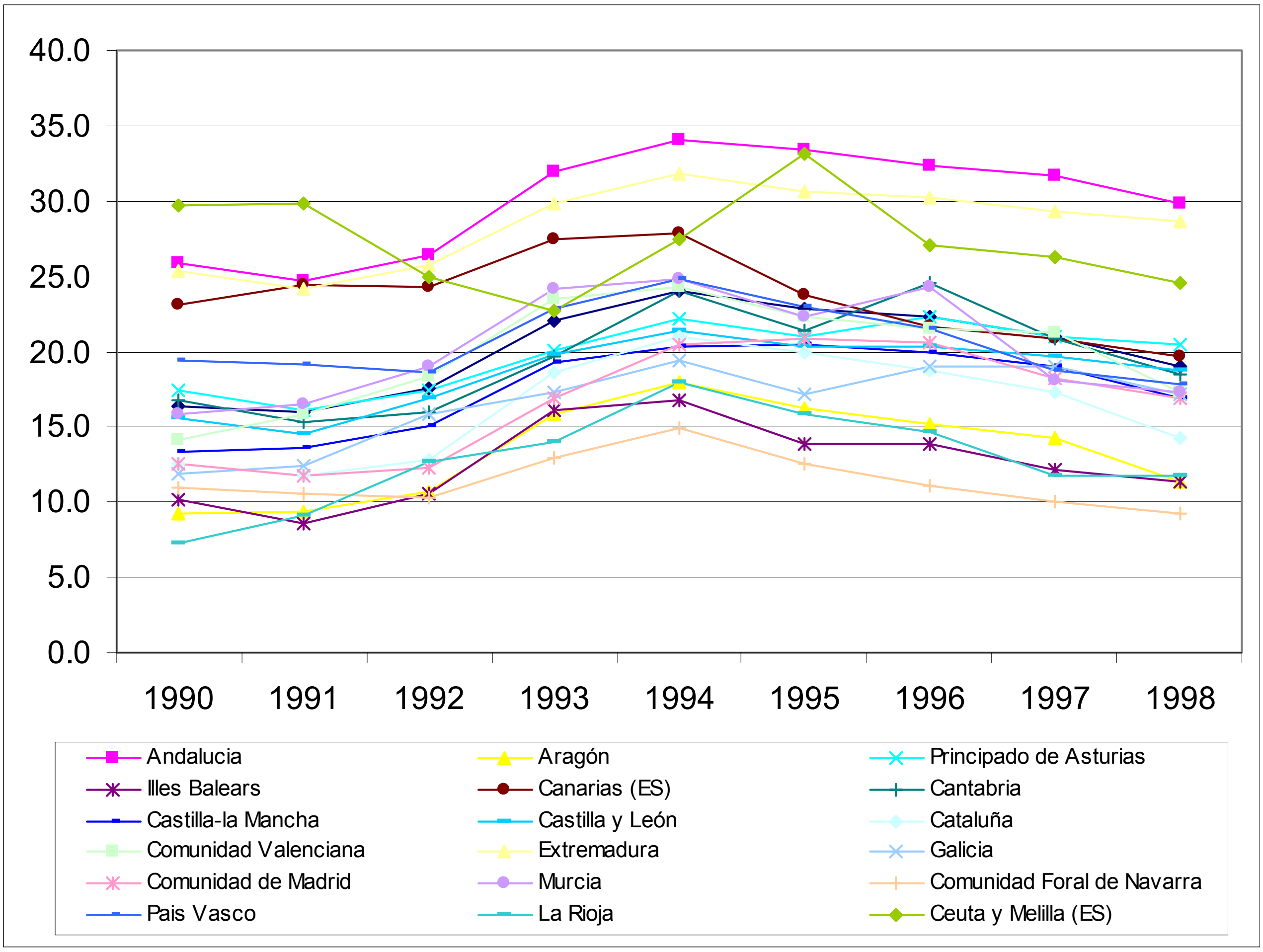


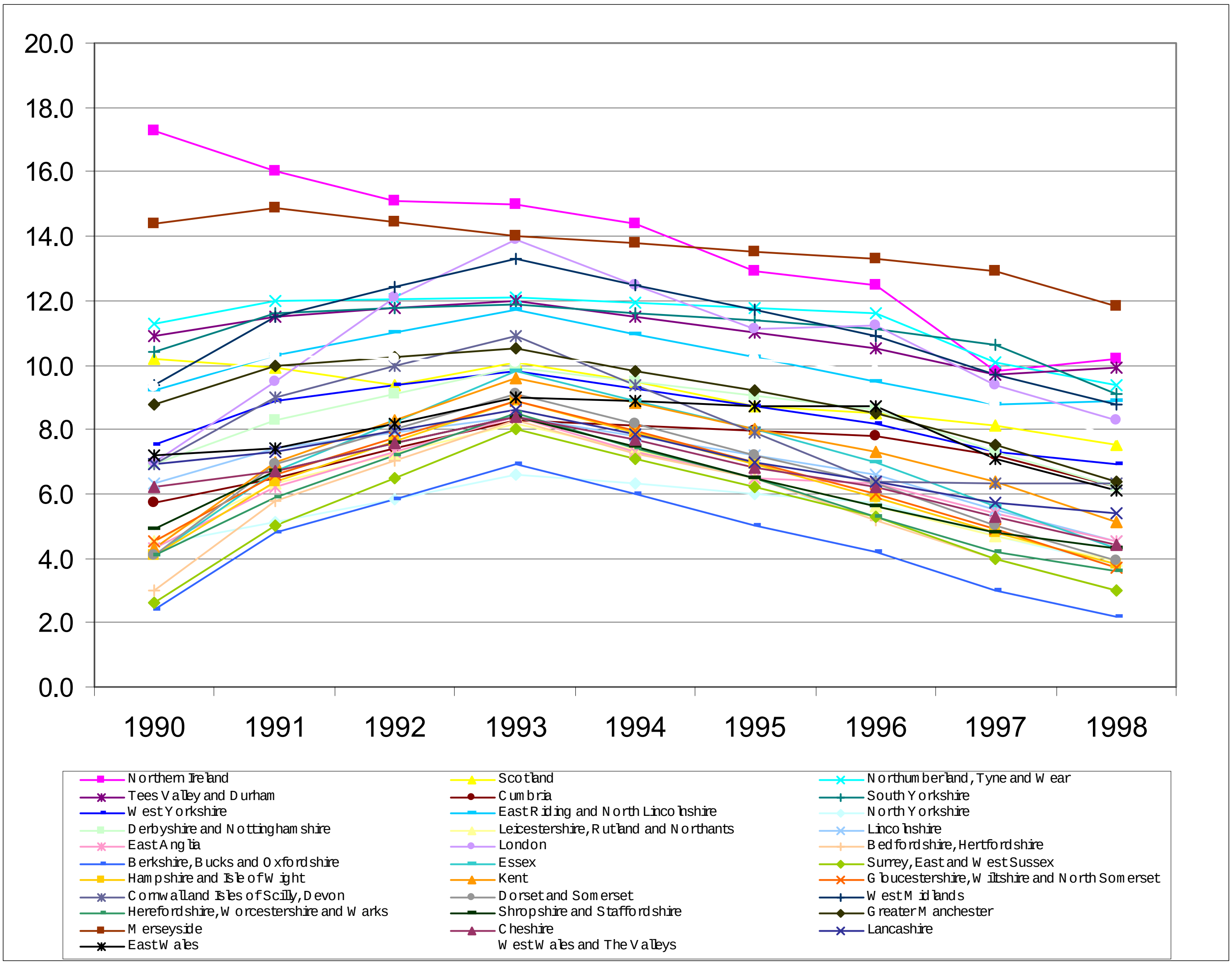
Figure 4.17: Unemployment rates, 1990-1998, Spain



4.4 Hypotheses

Each of the variables identified in the previous section is assumed to have an influence on migration. Hypotheses can be formulated that specify the nature of the relationships in more detail. However, we must be aware that relationships between migration and explanatory variables tend to be *time and scale specific*. In addition, relationships will vary by *age* and *other selective influences*, e.g. large population size may be important for young workers but not necessarily for young students. Moreover, we should not be surprised to find that *signs* of regression model parameters as well as their *significance,* will vary from place to place. Therefore, our general hypotheses for out-migration and for destination choice, presented in Tables 4.4 and 4.5 respectively, are somewhat tentative.

Figure 4.18: Unemployment rates, 1990-1998, United Kingdom



The hypotheses for GDP and unemployment indicate that for certain age groups or circumstances, the general hypothesis may not be applicable. The relationship between migration and unemployment in particular may depend on context. Whilst high or increasing unemployment may motivate people to move away in search of or to take up jobs elsewhere, these circumstances may result in declining out-migration because of the reduction in finance that allows people to migrate. In other words, the overall level of national prosperity or the regional prosperity relative to the national average may be important.

Hypotheses for destination choice mirror those for out-migration.

Table 4.4: Out-migration hypotheses

Variable	Sign(all ages)	Hypothesis
Population	Negative	Large areas with large populations (more jobs, services, etc) will have lower out-migration rates
Density	Negative	Higher densities will lead to lower out-migration rates
Immigration	Positive	High immigration areas will have high internal out-migration rates ('white flight' effect)
Accessibility	Negative	More accessible areas will have higher in-migration but lower out-migration rates
GDP	Negative	High income (growth) areas will have lower out-migration rates (*but* high income does increase propensity to move away at certain ages)
Unemployment	Positive	Higher (increasing) unemployment will lead to more out-migration (*but* high unemployment in a region relative to the nation may depress out-migration because the unemployed lack the resources for migration)
Employment	Negative	Higher employment (growth) will go together with lower out-migration rates
Housing	Negative	More housing construction will lower out-migration

Table 4.5: Destination choice hypotheses

Variable	Sign(all ages)	Hypothesis
Population	Positive	Larger destinations, more migration
Density	Positive	Higher density, more migration
Immigration	Negative	More immigrants, fewer internal migrants
Accessibility	Positive	Higher accessibility, more migration
GDP	Positive	Higher GDP attracts more migrants
Unemployment	Negative	Lower unemployment, more migration
Employment	Positive	More jobs, more migration
Housing	Positive	More housing construction, more migration
Distance	Negative	Further distance, fewer migrants
Contiguity	Positive	Contiguous areas have more migrants

4.5 Conclusions

There are important messages that come out of this chapter. Firstly, the construction of an explanatory migration model for use in different countries is reliant upon the provision of age-specific migration data that will have to be collected from NSOs because REGIO does not contain age-specific migration matrices at NUTS 2 level. It is possible to consider a solution to this that involves modelling aggregate flows from origins to destinations but this alternative is less optimal than the approach which models age-specific destination choice explicitly.

Secondly, the availability of data for non-demographic explanatory variables consistent between regions, across different countries and over time is a major constraint on the range of determinants that can be included in a generally applicable model. This is an important reason for selecting a relatively small set of variables but even then, it is likely that considerable estimation will be necessary to fill in gaps in the time series or adjustment will be required because of changes in the boundaries of NUTS 2 regions. The UK is a good example of a country where such estimation is necessary in part because of changes in regional definitions. There is also the need to consider the inclusion of variables that not only have some significance in a historical calibration but which will be appropriate for use in a projection context. Thus the conclusion is to opt for a relatively small number of explanatory variables, some of which may not be hugely significant but would be useful policy-related levers when applying the model for projection.

Thirdly, it is necessary to acknowledge the variation that exists in the number of spatial units, population size and magnitude of migration flows occurring in different EU countries. These are archetypal problems that confront cross-national comparative migration research in general and complicate the process of defining common measures of migrant behaviour or formulating generalised models of spatial patterns of migration.

Finally, whilst recognising that out-migration or destination choice propensities will depend upon selective characteristics such as age or sex (to a much lesser extent), it is useful to formulate a set of basic hypotheses that will serve as the reference point for the explanatory models that are considered in the subsequent part of the study.

5 Modelling internal migration

5.1 Introduction

The aim of our modelling exercise was to follow up on a number of recommendations for model improvement based on the previous EUROPOP regional projections. In particular, we focus on two new developments:

1. the introduction of explanatory variables in models of internal migration; and
2. the use of a life cycle approach, i.e. to study explicitly if different models apply to different ages.

Our analysis was based on data for Sweden, the Netherlands and the United Kingdom. Spain was used in order to see if the best common model was suitable for the Spanish case. The strategy was to model the internal migration processes for the period 1991-1995, and to validate the models by predicting the flows in 1996-1998 using the estimated model coefficients for each of the countries.

We model the internal migration process in two steps: first the out-migration rates, and second, conditional on out-migration, the probabilities of in-migrating to destinations:

$$m_{ij}^{as}(t) = m_i^{as}(t)\, p_{j|i}^{as}(t) \qquad (5.1)$$

All rates m and probabilities p are age- and sex-specific (a,s). $m_{ij}(t)$ is the rate of migration from region i to region j, $m_i(t)$ is the total rate of out-migration from region i, and $p_{j|i}(t)$ is the conditional probability of choosing destination region j after out-migration from region i. We construct models separately for $m_i^{as}(t)$ and $p_{j|i}^{as}(t)$.

The modelling strategy for both sub-models was to estimate two types of models for each country: first a purely *demographic model*, which is a model of only age-, sex-, region-, as well as time-specific rates and probabilities; and second, a *'best' model for each country using explanatory variables*. A purely demographic model is usually a good description of reality, but contains no causal "drives" that might produce change over time in the migration process. Comparing the goodness of fit of these two models for each country gives insight into the relative descriptive or explanatory power of non-demographic variables *vis-à-vis* a pure demographic approach.

In addition we want to know how well these models are able to *predict* migration rates and probabilities. Explanatory models may be less good in describing the base period trends, but they may be better able to predict future trends based on additional non-demographic information. This assumption was tested by using the period 1991-1995 as the base period for estimating the model parameters, and using the years 1996-1998 for prediction. This prediction phase is a form of external validation of the models. We used observed values of exogenous variables to predict migration outcomes, thus leaving aside for the moment the (important) question how to predict these exogenous variables. Of course, three years is a short period for prediction, but the available time series do not allow longer time periods for validation.

In the next step the 'best' models for each country were compared, in order to assess to what degree the models are comparable, and to see how well the optimal model for country A fits country B. This step should answer the question of whether 'one size fits all' or whether tailor-made models are necessary for each country. The results of the sub-models will be discussed below, starting with the out-migration model, followed by the destination probability model.

As will become clear in the analyses in this chapter, for prediction purposes it is not enough to decompose the migration process into two components, as done in equation (1). In addition, we need to take into account an explicit projection of the overall migration level *m(t)*. Comparable to origin and destination patterns, intensities can be simply extrapolated by using univariate statistical time series models (the purely demographic model) or they can be linked to economic developments. In section 5.4, attention will be paid to relationships between overall internal migration intensities and various economic indicators.

5.2 Out-migration

For out-migration, we estimated models of the form:

$$m_i^{as}(t) = \exp\{Z_i^{as}(t)\} \tag{5.2}$$

Equation (2) is a lograte model, where $Z_i^{as}(t)$ is the linear predictor. The model is estimated as a Poisson regression model using GLIM, in the following form:

$$\log M_i^{as}(t) = \log Pop_i^{as}(t) + \sum_k \beta_{ik}^{as} X_{ik}(t) + \varepsilon_i^{as}(t) \tag{5.3}$$

where $M_i^{as}(t)$ is the observed outflow from region *i* of age group *a* and gender *s* at time *t*, and $X_{ik}(t)$, *k*=1,...,*K* is a set of explanatory variables pertaining to region *i* at time *t*. These may also involve lagged values of regional characteristics. In general, we have taken the observed values of economic variables of the previous year *t*-1. β_k^{as}, *k*=1,...,*K* is a set of coefficients to be estimated. These coefficients may be age- and sex-specific, although in many cases they are generic for all population categories. In that case, the *a* and *s* superscripts may be dropped. $\varepsilon_i^{as}(t)$ is a random error term, which is Poisson distributed. A main assumption of the Poisson regression model is that the variance of the dependent variable is equal to the mean. This assumption is often violated, which is called 'over-dispersion'. As a result of over-dispersion, the estimates are unbiased but the standard errors are underestimated. We correct for over-dispersion using a simple method in GLIM to divide the error variance by the mean deviance. We treat the standard errors of the estimates as rough indicators of the significance of the coefficients.

Table 5.1 gives an overview of the explanatory variables used in the analyses. The table shows that for each variable, more than one operationalization was available: contemporaneous as well as lagged values, level as well as change variables, and for GDP and unemployment the absolute level as well as relative levels denoting the difference between the regional level and the national level for a given year. In initial modelling efforts, additional housing and immigration variables (the size of the immigration flow) were taken into account. Housing variables turned out to be only available in very few countries, and the results of using the immigration variable turned out to be of little added value for further exploration. The immigration variable was therefore dropped.

Table 5.1: Explanatory variables used in the analyses

	variable	operationalizations	Mnemonic
1	Population size	logarithmic form	LPOP
2	Employment	logarithmic lagged rate of change (t-1,t)	LEMP LEMPLAG EMPD
3	Regional mass	log of (population + employment)	LMASS
4	Unemployment	% unemployment lagged rate of change (t-1,t) regional difference with national average (lagged)	UNEMP UNEMP(LAG) UNEMPD UNEMPZ(LAG)
5	Gross regional product	gross regional product per capita at purchasing power parity GDP lagged rate of change (t-1,t) regional difference with national average (lagged)	GDP GDPLAG GDPD GDPZ(LAG)
6	Accessibility	demographic potential log form	ACCES LACCES
7	Population density	population per km^2 logarithmic	DENS LDENS

The demographic data for the analyses for Sweden, the Netherlands and the United Kingdom consist of out-migration by region, age (6 groups, 0-14, 15-19, 20-29, 30-44, 45-59, 60+), sex (only for Sweden and the Netherlands), and year (1991-1995). This amounted to 480 observations for Sweden (8 NUTS 2 regions), 720 for the Netherlands (12 NUTS 2 regions), and 960 for the United Kingdom (32 NUTS 2 regions according to the 1995 classification; Scotland has been taken as 1 region only). The data have a pooled structure, with both a cross-sectional and a time series dimension (panel).

At first, the purely demographic model was estimated, using the following model specification:

Sweden and the Netherlands: `O.A.S+T`
UK: `O.A  +T`

This fairly simple representation of model specifications is standard in “GLIM language”, where main and interaction effects are distinguished. In models involving categorical variables, such as the model forms here, this specification has a one-to-one relationship with necessary information of marginal totals of the corresponding cross-classifications. For instance, the UK model O.A+T specifies that the full table O.A.T, which has 960 cells, can be sufficiently described by the smaller table O.A, which has 32x6 = 192 cells, plus five scaling factors pertaining to each of the years 1991-1995 (i.e. the parameters in the time dimension T). Here O refers to a set of origin-specific parameters, representing regional base differences in out-migration rates; likewise A refers to the set of parameters, representing age-differences in out-migration rates; A.S refers to age- and sex-specific parameters. T finally, refers to the time path of out-migration. The demographic models specify a separate age curve of out-migration rates in every region. In Sweden and the Netherlands, these curves are also different for the sexes. These curves have the same form in every year, except for one scale parameter describing the overall change in out-migration rates for each year. This corresponds to a model of the out-migration rate as follows: $m_i^{as}(t) = A_t B_i^{as}$, where A_t is an overall scaling factor for each year, and B_i^{as} an age-, sex-, and region-specific base rate (for the UK there is no s index). Although more complicated models can be formulated, with region- and age/sex-specific scaling factors this formulation comes close to what is commonly used in demographic projection practice.

In the following step, models with explanatory variables were estimated. The base model here included a set of origin-specific parameters, age-(and sex-)specific parameters and the overall time-trend, but no interaction between origin and age (and sex): O+A.S+T. Subsequently, explanatory variables were added to the models (see Table 5.1 for descriptions). The best measure to assess the difference in fit for the various models is the mean likelihood ratio test statistic (LR/df), which takes into account the degrees of freedom involved. In Tables 5.2 to 5.4, estimation results are given for Sweden, the Netherlands and the United Kingdom respectively. In choosing the ‘best’ explanatory models, not only was goodness of fit an important criterion, but also the signs and significance of the coefficients were taken into account. If an explanatory variable was shown to be age-specific (e.g. A.GDP) the age pattern of the coefficients was important. In particular, we expect that high GDP in a region and low unemployment will give *lower* out-migration probabilities (see hypotheses in chapter 4). We expect this to be especially true for the younger age groups 15-19 and 20-29.

Table 5.2: Estimation results for Sweden, 1991-1995

Sweden Model specification n=8x6x2x5=480	LR test statistic: Deviance	df	Mean LR	Remarks
Null	478379	479	998.7	constant only
O.A.S+T	5706	380	15.0	demographic model
OA+T	10459	428	24.4	without gender
O+A.S+T	17485	457	38.3	basic model
O+A.S+T+LPOP	13341	456	29.3	+ LPOP: substantial improvement
O+A+T+LPOP	18019	462	39.0	without gender
O+A.S+T+O.LPOP	9953	446	22.3	
O+A.S+T+A.LPOP	11941	451	26.5	
O+A.S+T+LPOP+LACCES	13333	455	29.3	+ LACCES: no improvement
O+A.S+T+LPOP+ACCES	13336	455	29.3	
O+A.S+T+LPOP+A.ACCES	13010	450	28.9	
O+A.S+T+LPOP+LDENS	12630	455	27.8	+ LDENS: some improvement
O+A.S+T+LPOP+DENS	11574	455	25.4	DENS better than LDENS
O+A.S+T+LPOP+A.DENS	5609	450	12.5	
O+A.S+T+LPOP+LEMPLAG+GDPLAG+UNEMPLAG	12011	453	26.5	+ economic variables: minor improvement
O+A.S+T+LPOP+EMPD+GDPD+UNEMPD	12963	453	28.6	
O+A.S+T+LPOP+LEMPLAG+GDPZ+UNEMPZ	12722	453	28.1	
O+A.S+T+LPOP+LEMPLAG+GDPZLAG+UNEMPZLAG	12004	453	26.5	
O+A.S+T+LPOP+A.(LEMPLAG+GDPLAG+UNEMPLAG)	5538	438	12.6	economic var. x age: substantial improvement
O+A.S+T+LPOP+A.(EMPD+GDPD+UNEMPD)	11632	438	26.6	
O+A.S+T+LPOP+A.(LEMPLAG+GDPZ+UNEMPZ)	6430	438	14.7	
O+A.S+T+LPOP+A.(LEMPLAG+GDPZLAG+UNEMPZLAG)	6084	438	13.9	
O+A.S+T+LPOP+A.(GDPLAG+UNEMPLAG)	6178	444	13.9	best economic var.: GDPLAG + UNEMPLAG
O+A.S+T+LPOP+A.(LEMPLAG+UNEMPLAG)	8959	444	20.2	
O+A.S+T+LPOP+A.(LEMPLAG+GDPLAG)	6770	444	15.2	
O+A.S+T+LPOP+A.(LEMPLAG+GDPLAG+UNEMPLAG) +A.DENS	4060	432	9.4	+ A.DENS: further improvement
O+A.S+T+LPOP+A.(GDPLAG+UNEMPLAG)+A.DENS	4273	438	9.8	without A.LEMPLAG: almost as good
O+A.S+T+LPOP+GDPLAG+UNEMPLAG+A.DENS	5403	448	12.1	remove eco. var. x age: worse model

Table 5.3: Estimation results for the Netherlands, out-migration 1991-1995

Netherlands Model specification n=12x6x2x5=720	LR test statistic: Deviance	df	Mean LR	Remarks
Null	704929	719	980.4	constant only
O.A.S+T	1805	572	3.2	demographic model
OA+T	6807	644	10.6	without gender
O+A.S+T	28640	693	41.3	basic model
O+A.S+T+LPOP	27598	692	39.9	+ LPOP: slight improvement
O+A+T+LPOP	31693	698	45.4	without gender
O+A.S+T+O.LPOP	16788	681	24.7	substantial improvement
O+A.S+T+A.LPOP	23949	687	34.9	A.LPOP less important than O.LPOP
O+A.S+T+LPOP+LACCES	27588	691	39.9	+ LACCES: no improvement
O+A.S+T+LPOP+ACCES	27597	691	39.9	idem
O+A.S+T+LPOP+A.LACCES	26793	686	39.1	A.LACCES: again no improvement
O+A.S+T+LPOP+LDENS	27593	691	39.9	+ LDENS: no improvement
O+A.S+T+LPOP+DENS	27549	691	39.9	idem
O+A.S+T+LPOP+A.DENS	13476	686	19.6	A.DENS: significant improvement
O+A.S+T+LPOP+LEMPLAG+GDPLAG+UNEMPLAG	27573	689	40.0	+ economic variables: signs contrary to expectations
O+A.S+T+LPOP+EMPD+GDPD+UNEMPD	27566	689	40.0	no improvement
O+A.S+T+LPOP+LEMPLAG+GDPZ+UNEMPZ	27068	689	39.3	idem
O+A.S+T+LPOP+LEMPLAG+GDPZLAG+UNEMPZLAG	27122	689	39.4	idem, but coefficients with expected signs
O+A.S+T+LPOP+A.(LEMPLAG+GDPLAG+UNEMPLAG)	18311	674	27.2	signs contrary to expectations
O+A.S+T+LPOP+A.(EMPD+GDPD+UNEMPD)	25885	674	38.4	no improvement
O+A.S+T+LPOP+A.(LEMPLAG+GDPZ+UNEMPZ)	18192	674	27.0	economic var. x age: substantial improvement
O+A.S+T+LPOP+A.(LEMPLAG+GDPZLAG+UNEMPZLAG)	20275	674	30.1	idem, although somewhat less
O+A.S+T+LPOP+A.(GDPZLAG+UNEMPZLAG)	21121	680	31.1	best economic var.: GDPZLAG + UNEMPZLAG
O+A.S+T+LPOP+A.(LEMPLAG+UNEMPZLAG)	23053	680	33.9	
O+A.S+T+LPOP+A.(LEMPLAG+GDPZLAG)	22929	680	33.7	
O+A.S+T+LPOP+A.(LEMPLAG+GDPZLAG+UNEMPZLAG)+ A.DENS	10242	668	15.3	+ A.DENS: significant improvement
O+A.S+T+LPOP+A.(GDPZLAG+UNEMPZLAG)+A.DENS	12775	674	19.0	remove LEMPLAG: idem, although somewhat less
O+A.S+T+LPOP+GDPZLAG+UNEMPZLAG+A.DENS	13330	684	19.5	remove eco-age interaction: almost as good
O+A.S+T+LPOP+A.(GDPLAG+UNEMPLAG)+A.DENS	11605	674	17.2	best model for Sweden

Table 5.4: Estimation results for the United Kingdom, out-migration 1991-1995

United Kingdom Model specification n=32x6x5=960	LR test statistic: Deviance	df	Mean LR	Remarks
Null	3878299	959	4044.1	constant only
O.A+T	31474	764	41.2	demographic model
O+A+T	145760	919	158.6	basic model
O+A+T+LPOP	118825	918	129.4	+ LPOP: substantial improvement
O+A+T+O.LPOP	80896	887	91.2	substantial improvement
O+A+T+A.LPOP	68217	913	74.7	A.LPOP more important than O.LPOP
O+A+T+LPOP+LACCES	118526	917	129.3	+ LACCES: no improvement
O+A+T+LPOP+ACCES	118292	917	129.0	Idem
O+A+T+LPOP+A.ACCES	116300	912	127.5	A.ACCES: no improvement
O+A+T+LPOP+LDENS	118784	917	129.5	+ LDENS: no improvement
O+A+T+LPOP+DENS	118625	917	129.4	Idem
O+A+T+LPOP+A.DENS	65623	912	72.0	A.DENS: substantial improvement
O+A+T+LPOP+GDPLAG+UNEMPLAG	118445	916	129.3	+ economic variables: no improvement
O+A+T+LPOP+GDPD+UNEMPD	118390	916	129.2	
O+A+T+LPOP+GDPZ+UNEMPZ	118686	916	129.6	
O+A+T+LPOP+GDPZLAG+UNEMPZLAG	118800	916	129.7	
O+A+T+LPOP+A.(GDPLAG+UNEMPLAG)	78107	906	86.2	economic var. x age: substantial improvement
O+A+T+LPOP+A.(GDPD+UNEMPD)	97329	906	107.4	
O+A+T+LPOP+A.(GDPZ+UNEMPZ)	56466	906	62.3	
O+A+T+LPOP+A.(GDPZLAG+UNEMPZLAG)	105870	906	116.9	
O+A+T+LPOP+A.(GDPLAG+UNEMPLAG)+A.DENS	45168	900	50.2	+ A.DENS: substantial further improvement
O+A+T+LPOP+A.(GDPZ+UNEMPZ)+A.DENS	49941	900	55.5	
O+A+T+LPOP+GDPZ+UNEMPZ+A.DENS	65420	910	71.9	remove eco-age interaction: worse model

From Tables 5.2 to 5.4 we may conclude that, for Sweden and the UK, the model analyses using explanatory variables resulted in a similar result, taking into account that for the UK we do not have the gender dimension. The best explanatory model in both countries include the following variables: (log) population, lagged GDP, lagged unemployment rate, and density (all age-specific), in addition to a term specific to each region, and in addition to an age- (and sex-) specific set of terms. The specification of the models may be presented as follows:

Sweden: `O+A.S+T+LPOP+A.GDPLAG +A.UNEMPLAG +A.DENS`
UK: `O+A   +T+LPOP+A.GDPLAG +A.UNEMPLAG +A.DENS`

Since the demographic models have more coefficients to be estimated, for an equally good model one would expect a proportionally better fit.

The best model with explanatory variables for the Netherlands was slightly different. Instead of lagged GDP and unemployment rate, the lagged regional differences with the national averages of GDP and unemployment were included in the model, both not age-specific:

Netherlands: `O+A.S+T+LPOP+  GDPZLAG+  UNEMPZLAG+A.DENS`

Before discussing the interpretation of these models, we first compare the results with the purely demographic model. Table 5.5 gives the goodness of fit in terms of the likelihood ratio test statistic for both the demographic as well as the best explanatory models for all three countries.

Table 5.5: Likelihood ratio test statistic results for out-migration models in three countries

	demographic model AO(S)+T			**'best' explanatory models**		
	LR test stat.	**d.f.**	**mean LR**	**LR test stat**	**d.f.**	**mean LR**
Sweden	5706	380	15.0	4272	438	9.75
UK	31474	764	41.2	45168	900	50.2
Netherlands	1805	572	3.2	13330	684	19.5

Although for Sweden and the UK the best explanatory model captures the same variables, the goodness of fit of these models compared to the best demographic models does not point to one overall conclusion. For Sweden the explanatory model gives a better fit to the data, whereas for the UK the reverse is true. Taking also the results of the Netherlands into account, we may conclude that in the Netherlands the demographic model gives an exceptionally good fit, when judged from the mean LR. This is an indication that in the Netherlands the structure of the out-migration process is relatively time-invariant.

For the UK and Sweden, the signs, significance and age pattern of the explanatory variables in the best models broadly fulfilled the requirements that high GDP and low unemployment will give lower out-migration probabilities, although the age patterns are not always exactly as expected. In

Tables 5.6 and 5.7 the coefficients, standard errors and t-values are given of the explanatory variables of the best economic models for Sweden and the United Kingdom.

Table 5.6: Coefficients of best economic model of out-migration, Sweden

Best model	**Sweden (+ sex)**		
	estimate	**s.e.**	**t value**
LPOP	-0.316	0.138	-2.289
A1.GDPLAG	-0.048	0.029	-1.683
A2.GDPLAG	-0.063	0.033	-1.894
A3.GDPLAG	-0.075	0.027	-2.789
A4.GDPLAG	-0.091	0.028	-3.238
A5.GDPLAG	-0.091	0.032	-2.851
A6.GDPLAG	-0.111	0.035	-3.206
A1.UNEMPLAG	0.019	0.017	1.116
A2.UNEMPLAG	0.024	0.017	1.379
A3.UNEMPLAG	0.048	0.016	2.900
A4.UNEMPLAG	0.029	0.017	1.717
A5.UNEMPLAG	0.009	0.017	0.518
A6.UNEMPLAG	0.007	0.017	0.376
A1.DENS	-0.026	0.003	-9.101
A2.DENS	-0.027	0.003	-9.447
A3.DENS	-0.026	0.003	-9.444
A4.DENS	-0.025	0.003	-8.760
A5.DENS	-0.024	0.003	-8.462
A6.DENS	-0.022	0.003	-7.658

Note: Age groups are as follows: A1: 0-14, A2: 15-19, A3: 20-29, A4: 30-44, A5: 45-59, A6: 60+

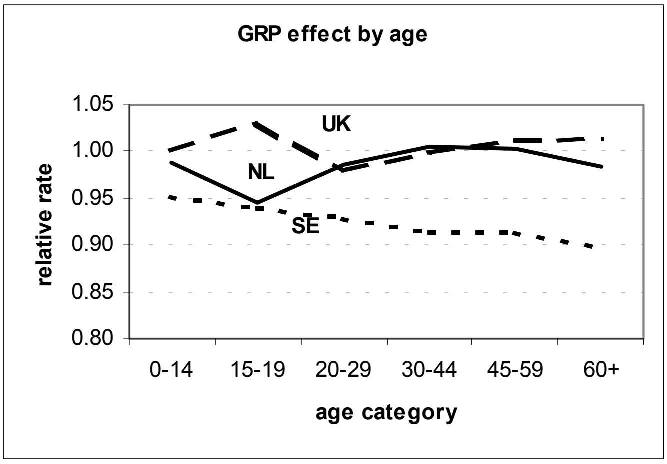
Table 5.7: Coefficients of best economic model of out-migration, United Kingdom

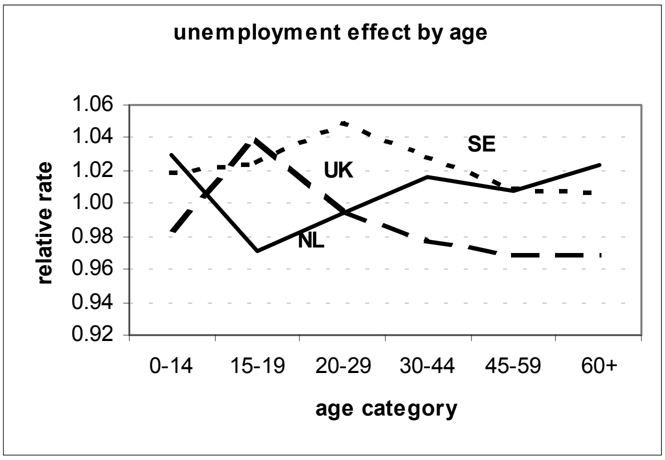
Best model	**United Kingdom (- sex)**		
	estimate	**s.e.**	**t value**
LPOP	-0.134	0.057	-2.341
A1.GDPLAG	0.002	0.015	0.118
A2.GDPLAG	0.031	0.015	2.100
A3.GDPLAG	-0.022	0.014	-1.504
A4.GDPLAG	-0.001	0.014	-0.041
A5.GDPLAG	0.012	0.015	0.820
A6.GDPLAG	0.014	0.015	0.950
A1.UNEMPLAG	-0.016	0.005	-2.951
A2.UNEMPLAG	0.039	0.006	6.663
A3.UNEMPLAG	-0.005	0.005	-0.959
A4.UNEMPLAG	-0.023	0.005	-4.470
A5.UNEMPLAG	-0.032	0.006	-5.117
A6.UNEMPLAG	-0.032	0.006	-5.129
A1.DENS	0.000	0.000	0.074
A2.DENS	0.000	0.000	-0.508
A3.DENS	0.000	0.000	-0.032
A4.DENS	0.000	0.000	0.231
A5.DENS	0.000	0.000	0.225
A6.DENS	0.000	0.000	0.420

Note: Age groups are as follows: A1: 0-14, A2: 15-19, A3: 20-29, A4: 30-44, A5: 45-59, A6: 60+

Figure 5.1 shows the coefficient values of GDP in the form of relative rates (multiplication factors of a base rate) for each country. In the UK, the effect of GDP is smaller than 1 for the 20-29 year age group, although larger than 1 for 15-19, and zero for the other categories. The positive effect for the age group 15-19 might indicate a welfare effect, which enables students to leave home and move to university cities outside their region of residence. For Sweden, the relative rate is smaller than 1 throughout, although less so for the younger age groups. Results for unemployment show that out-migration is higher in regions with high unemployment, for the 15-19 in the UK, and for the ages 15-44 in Sweden.

Figure 5.1: Age patterns of GDP and unemployment coefficients (relative rates) in the Netherlands (NL), Sweden (SE) and the United Kingdom (UK), 'best' model specification





The age patterns of the results of the SE-UK-'best' model (further specified as 'the common model') for the Netherlands are somewhat different (Table 5.8), although the fit is better than the model that we finally chose as the 'best' model for the Netherlands (a deviance of 11,605 with 674 df, with an average LR of 17.2). Out-migration rates are higher in regions with low GDP for ages 15-19 and 20-29, as we hypothesized earlier. At the same time, out-migration rates are lower in regions with high unemployment, especially for the 15-19, and this is not according to common expectations. For the Netherlands, the alternative specification of unemployment and GDP in relative difference terms from the national average give the expected results (Table 5.9). The coefficients are not age-specific, in other words they show similar values across all ages. The coefficient of GDPZLAG (the lagged value of the relative difference of the regional GDP when taking the national average as 100) is negative, and the coefficient of UNEMPZLAG, similarly defined in terms of the difference with the national average, is positive.

A remark has to be made here. The negative coefficients in the UK model (Table 5.7) can be interpreted in the following way, following the arguments of Boyle (1993): The unemployed are mainly concentrated in social housing in the high unemployment regions. It is difficult for social housing tenants to find equivalent housing in a destination region. So unemployment suppresses outmigration. An alternative interpretation in the Netherlands might be that there is a greater chance that the response to unemployment is a change of job and commuting journey rather than a move of the house and job (at least in the Randstad).

In the literature we also find an interpretation of unemployment as well as low spatial mobility as indicators of regional deprivation (e.g. Van Solinge *et al.*, 1998). One could also argue that low spatial mobility may lead to high unemployment. However, if we are searching for explanatory models with economic key indicators as triggers for internal migration behaviour, we have to stick to the neo-classical assumption of high out-migration following high unemployment.

Table 5.8: Coefficients of the common model of out-migration for the Netherlands

common model	Netherlands (+ sex)		
	estimate	s.e.	t value
LPOP	-0.367	0.062	-5.950
A1.GDPLAG	-0.013	0.018	-0.725
A2.GDPLAG	-0.056	0.018	-3.079
A3.GDPLAG	-0.015	0.017	-0.871
A4.GDPLAG	0.005	0.017	0.287
A5.GDPLAG	0.003	0.018	0.179
A6.GDPLAG	-0.016	0.019	-0.859
A1.UNEMPLAG	0.029	0.014	2.063
A2.UNEMPLAG	-0.029	0.015	-1.900
A3.UNEMPLAG	-0.006	0.012	-0.484
A4.UNEMPLAG	0.016	0.013	1.237
A5.UNEMPLAG	0.008	0.017	0.490
A6.UNEMPLAG	0.023	0.017	1.322
A1.DENS	0.002	0.001	1.435
A2.DENS	0.001	0.001	0.971
A3.DENS	0.001	0.001	1.037
A4.DENS	0.002	0.001	1.454
A5.DENS	0.002	0.001	1.500
A6.DENS	0.002	0.001	1.545

Note: Age groups are as follows: A1: 0-14, A2: 15-19, A3: 20-29, A4: 30-44, A5: 45-59, A6: 60+

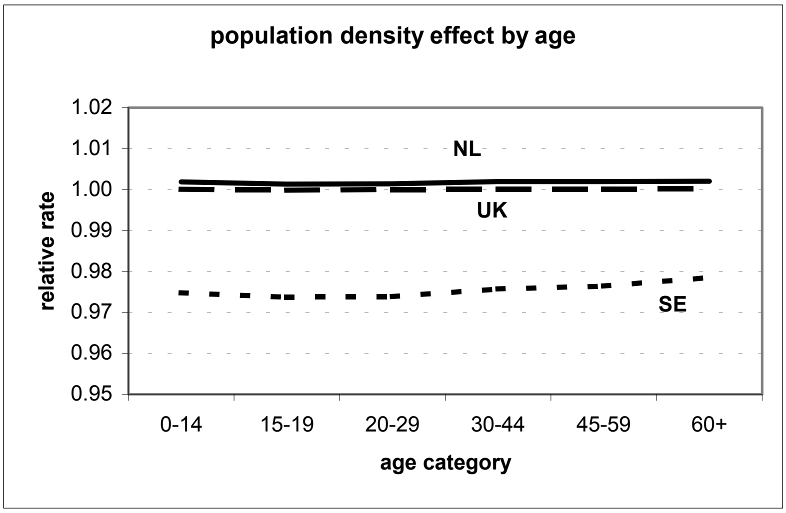
Table 5.9: Coefficients of best economic model of out-migration, the Netherlands

Best model	Netherlands (+ sex)		
	estimate	s.e.	t value
LPOP	-0.356	0.061	-5.883
GDPZLAG	-0.001	0.000	-2.315
UNEMPZLAG	0.001	0.000	2.682
A1.DENS	0.002	0.001	1.475
A2.DENS	0.001	0.001	0.937
A3.DENS	0.001	0.001	1.083
A4.DENS	0.002	0.001	1.540
A5.DENS	0.002	0.001	1.595
A6.DENS	0.002	0.001	1.587

Note: Age groups are as follows: A1: 0-14, A2: 15-19, A3: 20-29, A4: 30-44, A5: 45-59, A6: 60+

The effect of population density is very similar for the UK and the Netherlands, but very different for Sweden (Figure 5.2). For the Netherlands the results of the common model and the 'best' NL-specification are highly similar for this variable. In both the UK and the Netherlands the relative rates are larger than 1, implying that higher density leads to higher out-migration, for all ages, although more so for older ages. In Sweden, the coefficients are highly significant and the sign is different: higher density leads to *lower* out-migration. In the Netherlands and the UK negative effects of urban density prevail, especially for the older, whereas in Sweden the positive effects of higher densities prevail. This difference in result is not unexpected given the total different layout of the urban system in the countries involved.

Figure 5.2: Age patterns of population density (relative rates) in Sweden (SE), the Netherlands (NL) and the United Kingdom (UK), 'best model specification



The results of the other terms may be summarized as follows:

LPOP: the larger the region, the lower the probability of out-migration; the smaller the region, the larger the probability that a move will result in crossing the region border, and hence result in an out-migration. The coefficient is indeed negative in all countries, and broadly similar.

Age and sex: Figure 5.3 gives the age curves for three countries. These curves are as expected with high relative migration rates for the young and low for the old.

Origin-specific rates: there is a large variation across regions in region-specific effects, which cannot be interpreted easily, without looking simultaneously at the other included effects. Note that these base rates reflect differences between regions that cannot be explained by the other variables.

We also estimated models without the sex dimension, in order to compare directly with the UK model. The model fit is of course worse, but not dramatically so. For the Netherlands, the increase in LR is about 35 percent, and in Sweden about 50 percent.

The time trend is quite different between the countries, as depicted in Figure 5.4. For the UK it is upward, for Sweden it is initially downward, and later up, and for the Netherlands there is hardly any time trend apparent.

Figure 5.3: Age curves of base out-migration rates (relative rates) for three countries and two sexes; NL: Netherlands; SE: Sweden; UK: United Kingdom; m: males; f: females; t: total

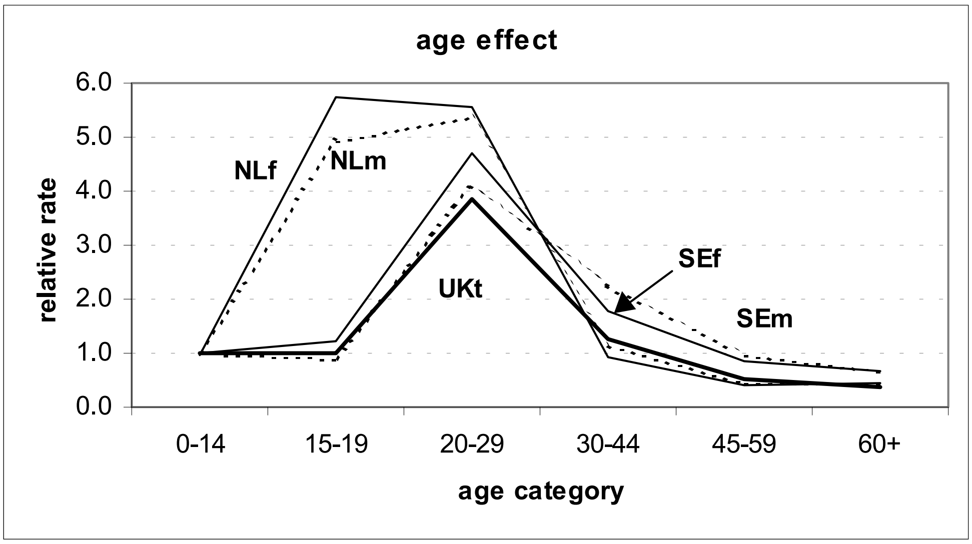
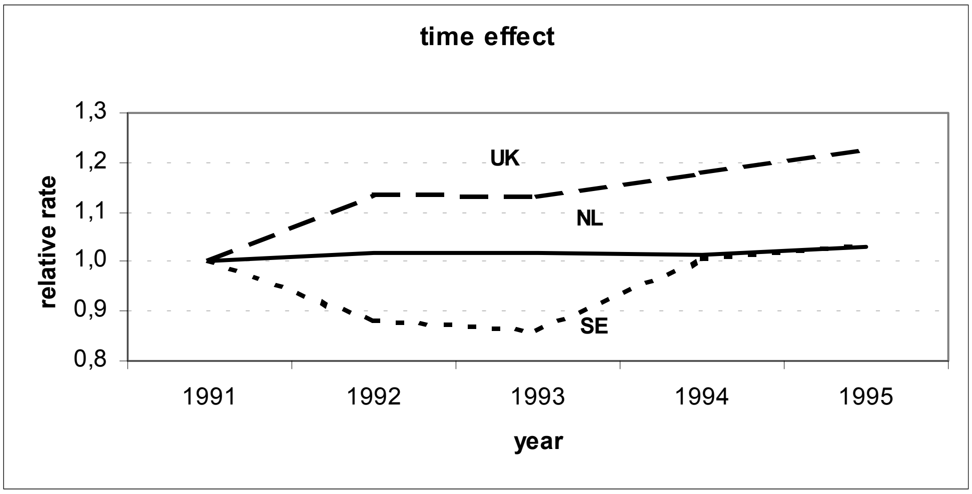


Figure 5.4: Time effect (relative rates) for three countries (1991 = 1.0).



We may conclude from the model fitting exercise using explanatory (non-demographic) variables that in the three countries involved, it is almost true that ‘one size fits all’, although there remain a number of important differences. First, in the Netherlands we used other indicators of GDP and unemployment: lagged values of the regional difference compared to the national value, whereas in the UK and Sweden we used lagged values of GDP and regional unemployment rates exactly. In a cross-sectional model, this would only affect the model intercept, not the coefficient, but in a pooled model such as used here, the change over time is important as well. Second, the values of the coefficients, and the age pattern of the coefficients are not always similar. This restricts the general applicability of one and the same model. Third, we did not explore the gender dimension in the UK case, and therefore we do not know if the same sex-pattern holds here as well.

Apart from model fit we also tested the models for their predictive power in the period 1996-1998. Figures 5.5 to 5.7 show the results for the three countries and two models. A perfect prediction would give data points on the diagonal line (expected = observed). If the data cloud is centred on the diagonal, the prediction is not biased, but may be not very precise, if the points do not lie close around the diagonal. If the data points lie around a line that deviates from the diagonal, there is a systematic bias in the prediction. Apart from this bias, the prediction may be precise or imprecise. When regressing the predicted rates against the observed rates, the bias is given by the deviation of the regression slope from 1. Values larger than 1 indicate systematic over-prediction, and values below 1 systematic under-prediction. The R^2 is an indicator of the precision of the prediction.

From Figures 5.5 to 5.7, we may conclude that for Sweden and the UK out-migration is systematically under-predicted, while for the Netherlands predicted values for both models are very close to the observed ones. In general, the demographic model outperforms the economic model.
Finally we looked at the Spanish case. Note, however, that for Spain we only have data for 1994 not distinguished by sex and for a different age distribution (7 age groups: 0-15, 16-24, 25-34, 35-44, 45-54, 55-64, 65+). First, we estimated the common model that gives a reasonable good fit in all other three countries. Since we have only one year at our disposal for Spain, we can only estimate a cross-sectional model without time dimension. The common model estimated in the other countries is a saturated model in the Spanish cross-sectional case, i.e. there are as many parameters as there are observations. This gives by definition a perfect fit, so we cannot judge the validity of the model. Results show that the following variables are important:

- LPOP (Population size: coefficient: -0.55, t-value –3.75): this is in line with previous results: the larger the region, the smaller the probability of out-migration;
- GDPLAG (lagged values of GDP per capita): negative values up to age 55, positive values for higher ages. This implies that for the economically active ages a higher GDP leads to lower out-migration. For the elderly the reverse is true. This is in line with expectations;
- for the age group 16-24, unemployment is almost significant, but with the wrong (-) sign; and
- the same holds for the same age group for density, with higher out-migration rates for lower density regions.

Although we used a saturated model, we applied both the estimated demographic and economic model in predictions for 1998. Results are shown in Figure 5.8. From this picture it is clear that there is again systematic bias in the prediction, which is larger for the economic model.

Figure 5.5: Predicted out-migration rates 1996-1998 for Sweden in two models

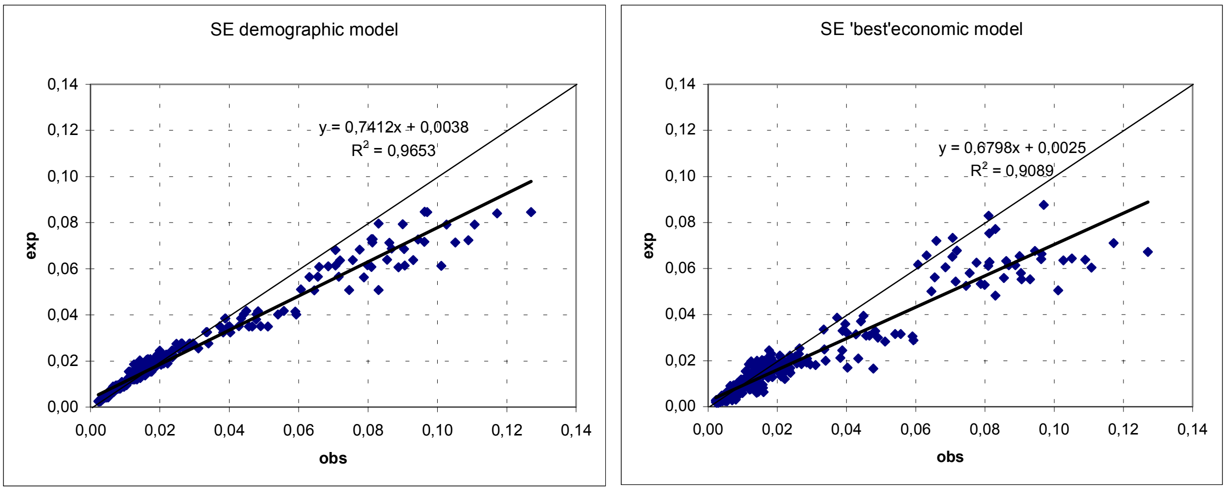


Figure 5.6: Predicted out-migration rates 1996-1998 for UK in two models

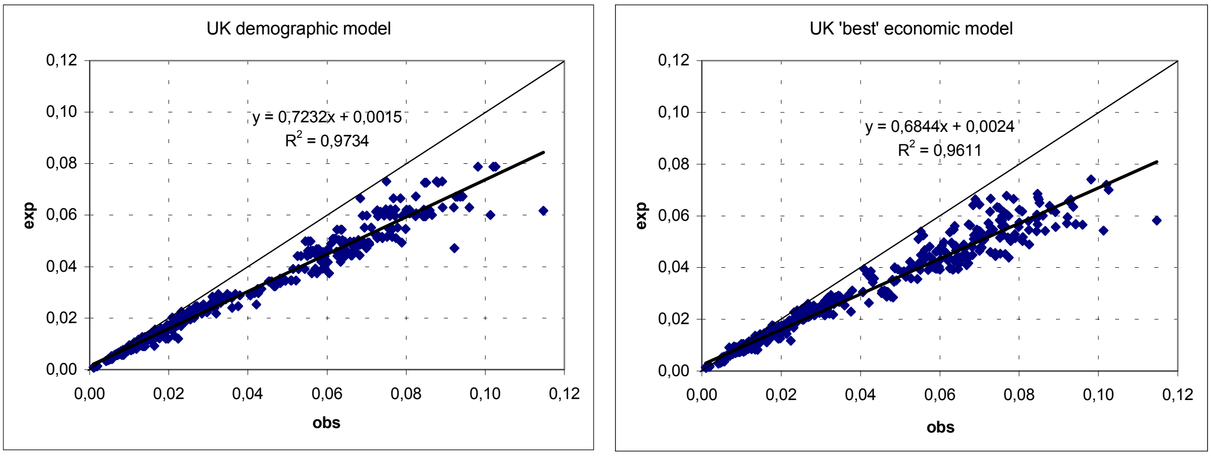


Figure 5.7: Predicted out-migration rates 1996-1998 for the Netherlands in two models

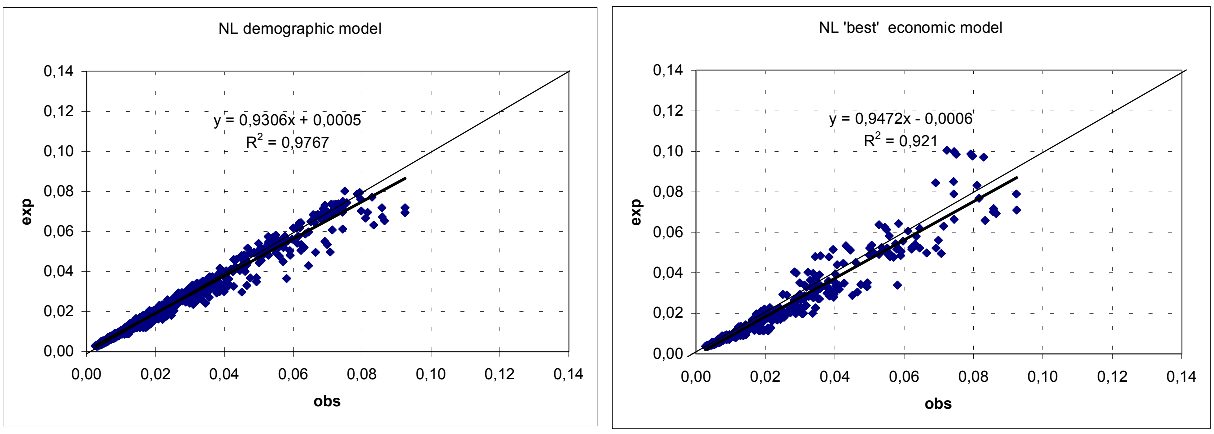
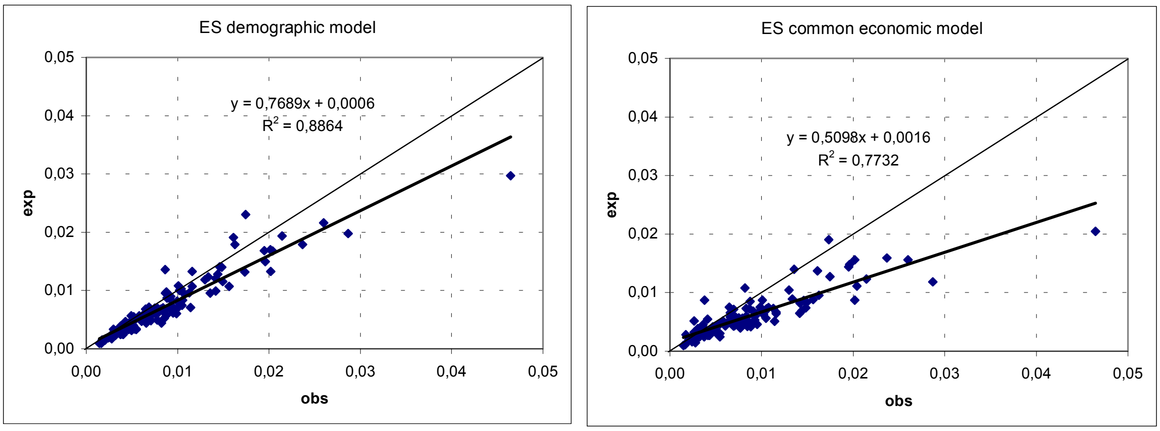


Figure 5.8: Predicted out-migration rates 1998 for Spain in two models



5.3 The destination choice sub-model

For the destination choice sub-model, we modelled the flow from *i* to *j*, conditional on leaving *i*. This flow is determined by two sets of factors: attractiveness factors of the region of destination *j*, and the friction of distance between *i* and *j*. For both dimensions, we have two options. First, the attractiveness of the destination may be modelled in a purely demographic fashion, *or* by using exogenous explanatory variables. Second, the distance decay between *i* and *j* can be modelled in a demographic framework using observed historical flows *or* by using an explicit distance function. Demographers generally use the method of historical migration matrices when taking observed flows of previous time periods for prediction purposes. The combination of a demographic approach for the attractiveness function and a non-demographic approach of a function for the distance decay is not very useful, therefore in total there are three model options, as given in the following scheme:

	Attractiveness function	
Distance function	demographic	exogenous information
Demographic	1	2
spatial interaction function	x	3

For each country we will model these three model types.The model form is a multinomial logit, which has the following form:

Model 1: Demographic model with OD distance function

$$p_{j|i}^{as}(t) = \frac{\overline{M_{ij}} A_j^{as}}{\sum_{k=1}^{I} \overline{M_{ik}} A_k^{as}} \qquad (5.4)$$

where $\overline{M_{ij}}$ is the historical flow from *i* to *j* (e.g. the average of the last five years), and A_j^{as} is an attractiveness factor for region *j*, which may be age- and sex-specific.

This model merely says that out-migration flows out of *i* are distributed over the destination regions according to historical destination shares, adjusted by destination specific constants. In the simplest approach, the *A*'s are set to 1.

Technically this model is estimated as a Poisson model of the flows $M_{ij}^{as}(t)$ in GLIM using the following specification:

$$\log M_{ij}^{as}(t) = \mu_i^{as}(t) + \log \overline{M_{ij}} + \alpha_j^{as} \tag{5.5}$$

where $\log \overline{M_{ij}}$ is called an *offset*, viz. a variable with an a priori fixed coefficient value, which remains outside of the estimation procedure. It is therefore subtracted from the dependent variable *before* the estimation procedure; hence the name *offset*. The α values are a set of coefficients, one for each (*a,s,j*) combination, and the $\mu_i^{as}(t)$ are proportional to the log of the observed out-migration flows out of *i*. Equation (5.5) is transformed into (5.4) as follows:

$$p_{j|i}^{as}(t) = \frac{M_{ij}^{as}(t)}{\sum_{k=1}^{I} M_{ik}^{as}(t)} = \frac{\exp\{\mu_i^{as}(t) + \log\overline{M_{ij}} + \alpha_j^{as}\}}{\sum_{k=1}^{I} \exp\{\mu_i^{as}(t) + \log\overline{M_{ik}} + \alpha_k^{as}\}} = \frac{\exp\{\log\overline{M_{ij}} + \alpha_j^{as}\}}{\sum_{k=1}^{I} \exp\{\log\overline{M_{ik}} + \alpha_k^{as}\}} = \frac{\overline{M_{ij}} \exp\{\alpha_j^{as}\}}{\sum_{k=1}^{I} \overline{M_{ik}} \exp\{\alpha_k^{as}\}} \tag{5.6}$$

where $A_j^{as} = \log \alpha_j^{as}$. Note that the parameters pertaining to the out-migration flows cancel out. A shorthand notation of this model is:

```
O.A.S.T + D.A.S + {OD}
```

where {OD} is the historical migration matrix, included as an offset.

Model 2: Explanatory model with OD distance function

$$p_{j|i}^{as}(t) = \frac{\overline{M_{ij}} \exp\{\alpha_j^{as} + X_j(t)\beta_i^{as}\}}{\sum_{k=1}^{I} \overline{M_{ik}} \exp\{\alpha_k^{as} + X_k(t)\beta_i^{as}\}} \tag{5.7}$$

Here we have added explanatory variables X_j to the demographic model, in order to explain the relative attractiveness of the destinations. The constants α_j are the intercepts of the linear predictor of the attractiveness function, and are comparable to log *A* in model 1 (eq. 5.5). The coefficients β may be age-, sex- and origin-dependent.

Equation (5.7) is estimated in GLIM using the following specification:

$$\log M_{ij}^{as}(t) = \mu_i^{as}(t) + \log \overline{M_{ij}} + \alpha_j^{as} + X_j(t)\beta_j^{as} \tag{5.8}$$

and the transformation of (5.8) to (5.7) is similar to (5.6). This model may be abbreviated to:

```
O.A.S.T + D.A.S + {OD} + A.S.XVARS
```

where XVARS refers to the set of explanatory variables.

Model 3: Spatial interaction model:

The formula is:

$$p_{j|i}^{as}(t) = \frac{\exp\{\alpha_j^{as} + X_j(t)\beta_i^{as} + F_i^{as}(D_{ij}\}}{\sum_{k=1}^{I} \exp\{\alpha_k^{as} + X_k(t)\beta_i^{as} + F_i^{as}(D_{ik}\}\}} \tag{5.9}$$

Here, the historical flow matrix is replaced by a spatial interaction function $F_i^{as}(D_{ij})$ of distance $D_{ij.}$ The function may be origin-, age- and sex-specific. In our applications we use straight line distances between the centroids of the regions, and a contiguity indicator, which is equal to 1 if two regions border each other, and 0 otherwise. The implementation in GLIM is similar to equation (5.8). This model may be abbreviated to:

O.A.S.T + D.A.S + O.A.S.DVARS + O.A.S.XVARS

where DVARS refers to the distance variables LDIST (straight line distances, logarithmic) and CONT (contiguity).

These three models were estimated for Sweden, the Netherlands and the UK. As with the out-migration model, we fit optimal models per country, and compare the results across countries. The ‘best’ models are then applied to Spain.

The X-variables tested in the destination choice models were a regional mass indicator (population plus employment summed: LMASS), unemployment, gross regional product GDP, accessibility and population density (see operationalisations in Table 5.1).

Table 5.10 gives goodness-of-fit statistics of the three models for Sweden, the United Kingdom and the Netherlands. The exact specification of the economic models e described below.

Table 5.10: Goodness of fit of three destination choice models for three countries

	Demographic model			Economic model +OD			Economic model plus spatial interaction function		
	LR	d.f.	Mean LR	LR	d.f.	mean LR	LR	d.f.	mean LR
Sweden	7629	2838	2.68	7542	2831	2.66	16901	2850	5.93
UK	175777	28614	6.14	175389	28608	6.13	483110	28539	16.9
Netherlands:	18749	7134	2.63	18714	7128	2.63	53803	7157	7.58

In the economic models, there are basically two options: either to include an intercept for each destination (the α_j terms) or not. If they are included, the only explanatory attractiveness variable that adds something to the explanation and has interpretable coefficients, is GDP lagged. The gain in fit, however, is only small. For instance, including all economic variables in the economic model + OD improves the fitted value with less than 10 percent in all three countries. Excluding these intercepts results in a large loss in model fit in all countries. On the other hand, in models without an intercept α_j it is possible to include more economic variables with significant coefficients. This could be interpreted as having better explanatory handles, but the attractiveness function may be misspecified. This is because the intercept captures all the effects of not-included variables that may explain systematic and persistent differences between regions. As an example we compare the model results for the Netherlands using two specifications for the attractiveness function, based on the model with distance and contiguity included. The two models are:

Model 3a: `O.A.S.T + A.D + O.LDIST + O.CONT + A.GPRLAG + LMASSA`
Model 3b: `O.A.S.T + O.LDIST + O.CONT + A.GDPLAG + A.LACCES + A.LDENS + LMASSA`

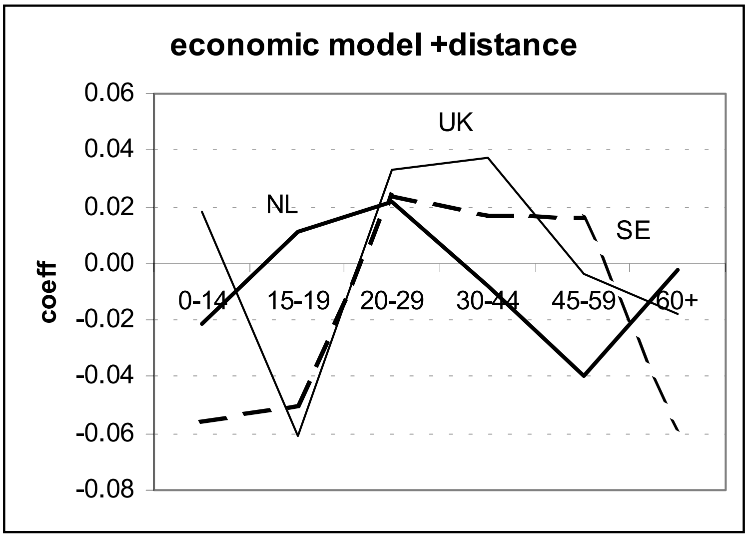
Note that the distance function is origin-specific. In model 3b the effect of the economic variables is age-specific. The coefficients of the economic variables are shown in Table 5.11. They are in general interpretable. In particular, GDP lagged is positive, and highest for the 20-29 group (age group 3). The spatial interaction variables (not shown here) are origin dependent. Distance is negative, and contiguity is mostly positive, indicating that neighbouring regions have a higher interaction than expected on the basis of distance alone. Accessibility (model 3b) is positive, except for the age group 15-19: more accessible regions are more attractive to immigrants. Density has a negative effect. It could be argued that this should be different for the younger age groups 15-29. The large negative value for this age group could be a sign of linear dependence among the explanatory variables.
The LR test statistic of model 3a is 53803 with 7103 degrees of freedom. Model 3b has a LR test statistic of 76531, with 7157 df, which is much worse. We prefer the model with intercept and GDP, which is simpler and gives much better fit.

The model with age-specific intercept and GDP is a likely candidate for the economic model for all three countries. Additional variables do not add much to the model fit, and coefficients have the wrong sign or are insignificant.

Figure 5.9 shows the age pattern of the coefficient of GDP. The expectation is that it is positive. This is in all three countries only true for part of the age range. In the Netherlands the age groups 15-29 are attracted to regions with high GDP. In the UK and Sweden this applies to a larger age category: in the UK in the age range 20-44 and in Sweden even until the age group 45-59. The result for the economic model with OD interaction term is broadly similar.

From this coefficient comparison it is clear that there is no general applicable age pattern of the effect of GDP on regional attractiveness.

Figure 5.9: Values of GDP lagged in destination choice models for three countries.



The other coefficients are in line with expectations. The A.D term gives age profiles of in-migration by region. In general, adding gender does not change the model very much. Distance is negative, although the effect is different over the origins. The coefficient for contiguity is sometimes positive and sometimes negative, thus correcting for the prediction error of the distance effect in shorter distances.

Figures 5.10 to 5.12 show the predictive value of the models for the three countries for the period 1996-1998. The Swedish results show that the economic model using historical flow data shows the largest dispersion in prediction of the destination choice probabilities, whereas the economic model using distance is more centred around the diagonal line of perfect fit. For the UK the results show that the demographic model gives the best prediction, and the economic model plus distance function the worst. The results for the Netherlands show again that the economic model using distance gives best results in prediction, and the economic model plus historical flows the worst.

Table 5.11: Coefficient values of two models of destination choice, using economic variables for the Netherlands

	Model 3a includes intercept		Model 3b no intercept	
Parameter	estimate	s.e.	Estimate	s.e.
LMASSA	-0.323	0.212	0.654	0.008
A1.GDPLAG	-0.027	0.024	0.000	0.006
A2.GDPLAG	0.009	0.027	0.131	0.006
A3.GDPLAG	0.019	0.015	0.093	0.003
A4.GDPLAG	-0.013	0.020	0.033	0.004
A5.GDPLAG	-0.044	0.033	0.015	0.007
A6.GDPLAG	-0.007	0.036	0.005	0.008
A1.ACCES			0.151	0.045
A2.ACCES			-0.201	0.054
A3.ACCES			0.101	0.028
A4.ACCES			0.223	0.037
A5.ACCES			-0.082	0.062
A6.ACCES			0.177	0.065
A1.DENS			-0.447	0.037
A2.DENS			-0.436	0.038
A3.DENS			-97.410	0.023
A4.DENS			-0.244	0.030
A5.DENS			-0.618	0.048
A6.DENS			-0.543	0.053
LR statistic	53803		76531	
d.f	7103		7157	
Mean df	7.57		10.69	

Note: Age groups are as follows: A1: 0-14, A2: 15-19, A3: 20-29, A4: 30-44, A5: 45-59, A6: 60+

Subsequently, the three models were applied to Spain. However, model 2 (A.GDPLAG plus OD interactions) gives the wrong sign of the coefficient of GDPLAG (negative). In model 3 (economic variables plus distance function) the coefficients are positive for age groups 16-35. Figure 5.13 gives the result of applying model 1 and 3 to the Spanish case. The figures show the prediction of the 1998 pattern using the 1994 model results. The demographic model is the most accurate whereas the economic model using distance shows a large dispersion around the diagonal. This model shows a serious underestimation of the higher destination probabilities and an over-prediction of smaller probabilities. The bad performance of the economic model comes as no surprise, since the results of the estimation using the 1994 data showed that the economic variables were not significant.

Thus, for Sweden and the Netherlands the economic model using a distance function proves better in predicting the destination choice probabilities in 1996-1998 than the models using historical flow data. This is a striking result, because the descriptive power of this model for the period 1991-1995 was markedly less than the other models. For the UK and Spain the conclusion is that the demographic model performs better, and the economic model far worse. This is supported by the result that the best fitting economic model in Sweden, the UK and the Netherlands cannot be transferred to Spain. In terms of transferability, the demographic model is the best option.

Figure 5.10: Prediction of destination choice probabilities 1996-1998, Sweden

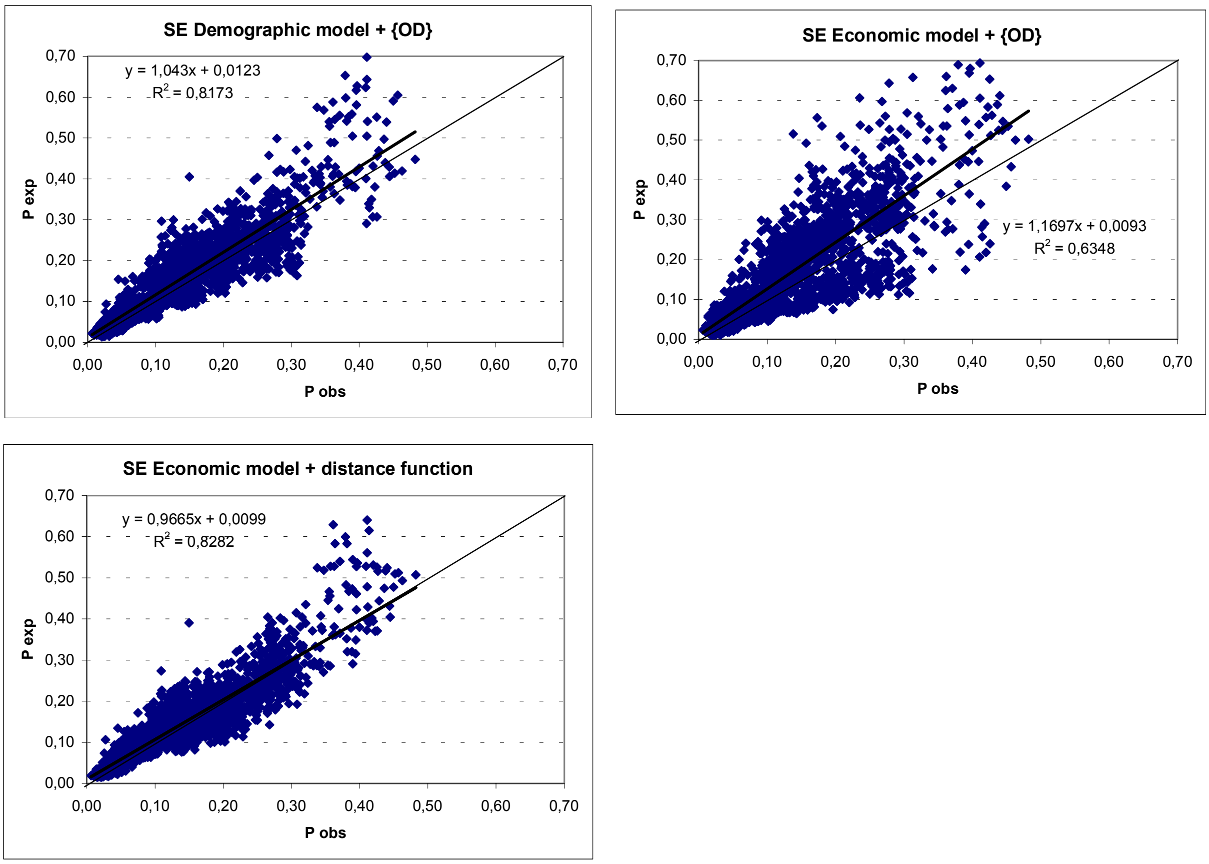


Figure 5.11: Prediction of destination choice probabilities 1996-1998, UK

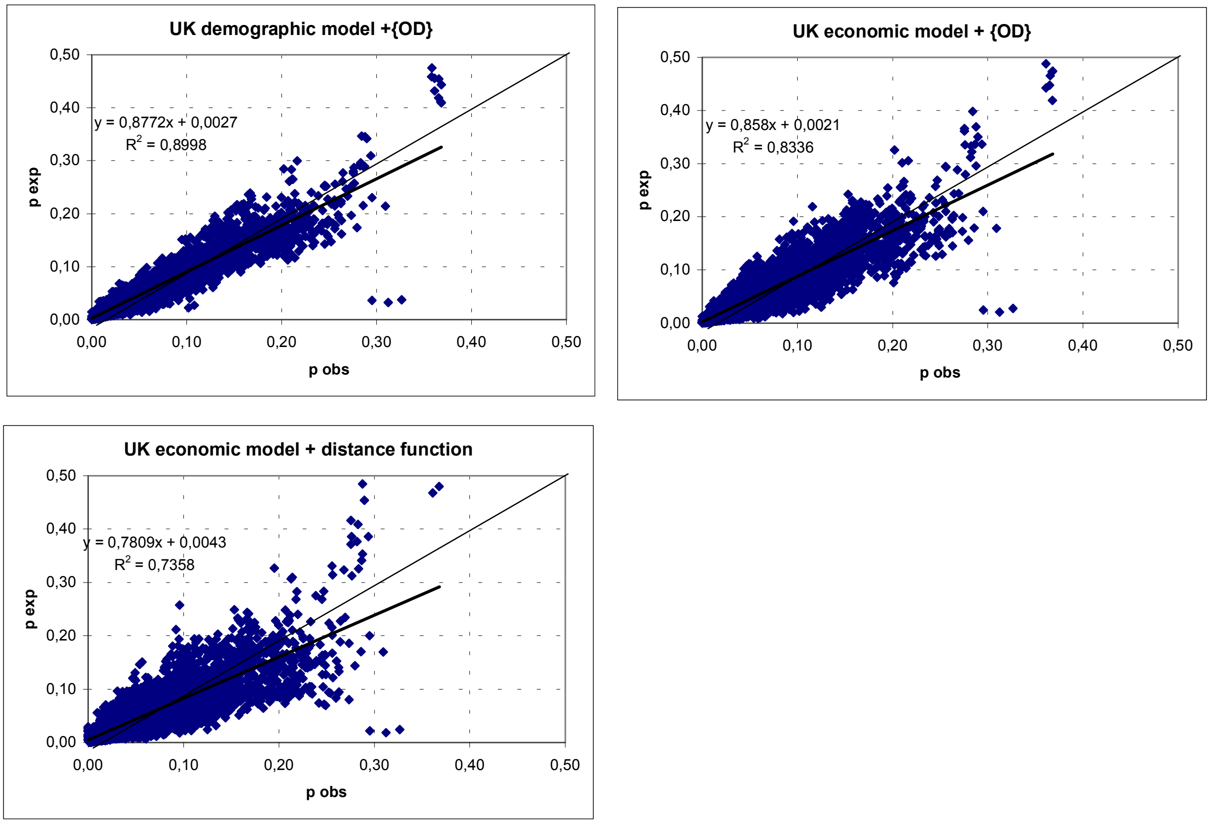


Figure 5.12: Prediction of destination choice probabilities 1996-1998, the Netherlands

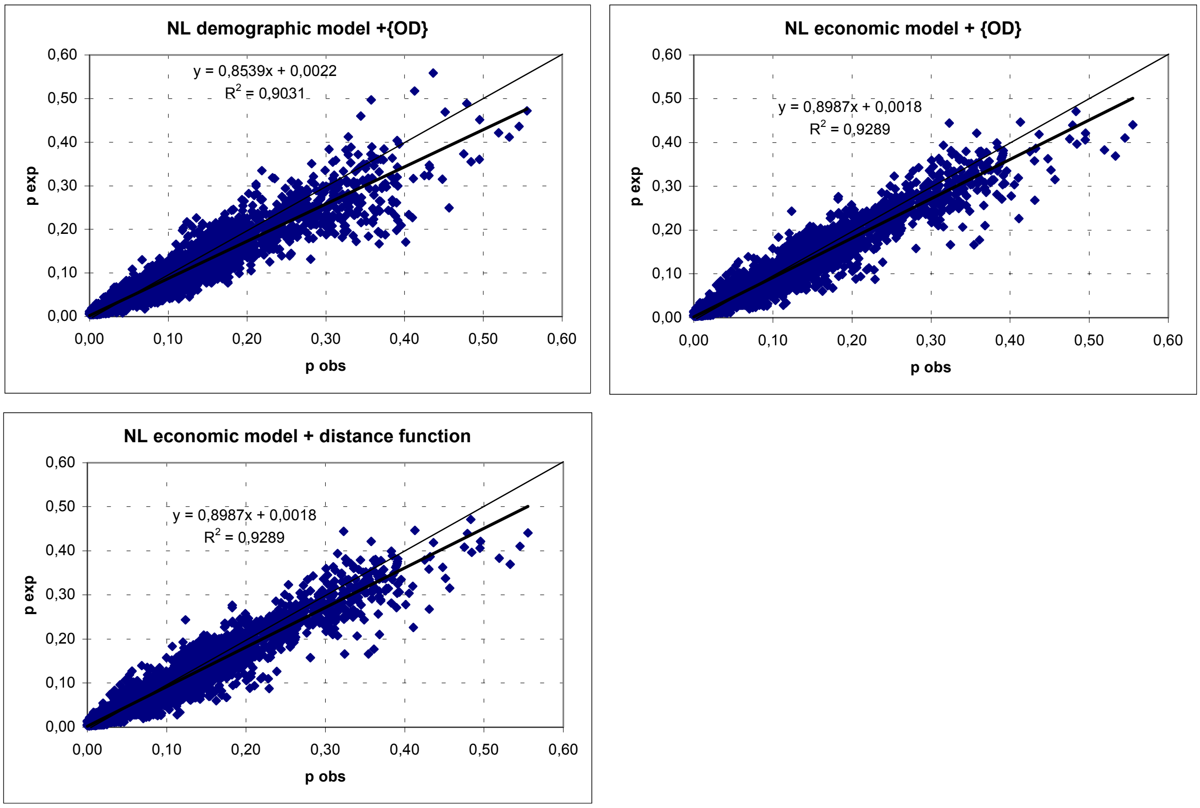
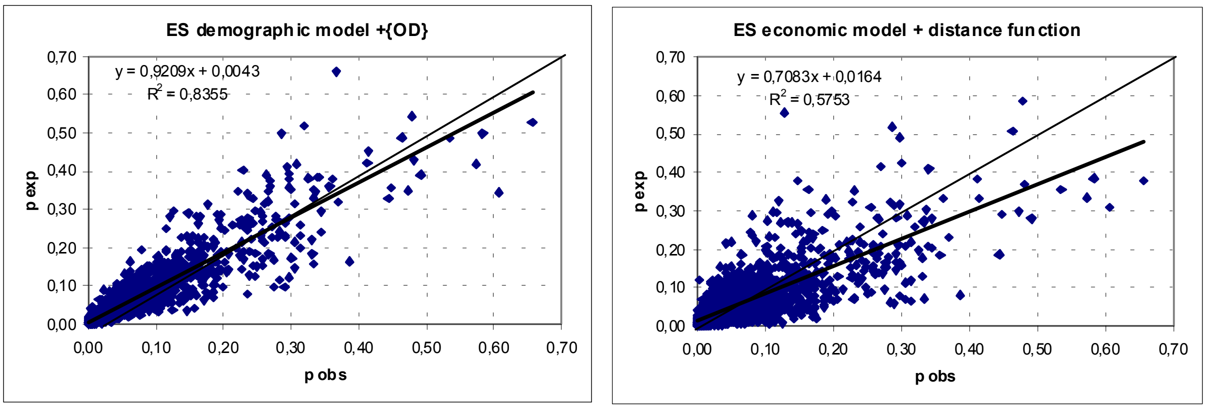


Figure 5.13: Prediction of destination probabilities 1998, Spain



5.4 Overall internal migration intensities [2]

In studying the relationship between internal migration intensities and economic variables, a slightly different approach has been used. Hypothesized relationships were estimated on a country-by-country basis. Regression models were estimated with crude migration rates (CMRs) used as dependent variables. Migration intensities were analysed for six European countries: the northern countries of Finland (1981-1998) and Sweden (1980-1999), the western countries of the Netherlands (1972-1998) and Switzerland (1981-1998), and the southern countries of Italy (1972-1995) and Spain (1979-1998). As sex-specific age curves (based on rates) appeared remarkably stable over time, it was decided to leave the general sex-specific age curves constant and to concentrate on developments in the overall migration level.

Hypotheses

Although there is only a limited number of empirical research which has analysed the relationship between national economic developments and migration intensities, a number of economic determinants have been mentioned in the literature which seem to be related to internal migration developments. These determinants cover both the economic business cycle, as well as more structural developments.

Regarding the relation between migration and the business cycle, it is assumed that the level of inter-regional migration should rise during times of prosperity and decline in periods of recession. Both housing market and labour market processes may contribute to this situation. In times of economic downturn, job and housing moves may be postponed, whilst during times of economic buoyancy, these intended moves may occur (Dunford and Fielding, 1997). In this respect, several studies have found a positive relationship between GDP and migration intensities (Birg, 1983; Milne, 1993; Kemper, 1997). In addition, relationships were found between employment and unemployment growth (Barff, 1990; Kemper, 1997; Öberg, 1997). Another important indicator is the mortgage rate of loans for private owned houses. In the Netherlands, for instance, interest on mortgages is tax deductable, and therefore housing market developments in the Netherlands may be influenced by the interest rate charged on mortgages (Van Fulpen, 1985; Van Wissen *et al.*, 1991).

As far as structural changes in the economy are concerned, nowadays a slowing down of inter-regional migration might be expected (Mønnesland, 1997). The change from mainly primary industries with widespread regional patterns, through manufacturing industries with an urban dominated locational pattern, to the service sector with a more local oriented pattern, has eventually led to more inter-regional stability. Personal services are often consumed locally and therefore, a larger part of local production depends on local demand. Consequently, a rise of the service sector will probably result in declining migration intensities (Bengtsson and Johansson,

[2] This section is largely based on a related study, entitled 'Economic determinants of regional migration', which is part of the research programme 'Towards a dynamic scenario model of economic determinants of European population dynamics', financially supported by the Netherlands Organisation for Scientific Research (NWO). Preliminary results of this study were presented at the European Population Conference 2001, Helsinki (Van der Gaag and Van Wissen, 2001).

1995; Öberg, 1997). At first, however, with the transition from agricultural to industrialised societies, internal migration was mainly characterised by flows from rural to urban areas. Another structural development contributing not only to declining migration levels, but also to a weakening of the relationship between the business cycle and migration, is the increase of female labour participation (Kemper, 1997; Rees *et al.,* 1998). As it is in general more difficult to find two new suitable jobs than just one, with the increase in two-earner households, commuting is probably increasing as alternative to migration. Finally, the ageing of the population will play a part as well. As the population in western Europe becomes older, there will be a tendency to lower inter-regional migration. Older workers are less inclined to change jobs, either because employers are prepared to pay more to keep older workers and their knowledge, or because older people are less attractive in the labour market because their knowledge has not kept up with modern technology. In both cases, however, ageing of the labour force will lower the level of job-related migration (Öberg, 1997).

Relationships have been studied between internal migration propensities and the following economic indicators: GDP per capita, unemployment, inflation, interest, labour force participation of women, employment, employment in services, and ageing of the labour force. For the economic business cycle indicators, different specifications were used in the models: (a) a contemporaneous relationship; (b) a lagged relationship, for instance by one year (**(-1)[3]); (c) a relationship based on differences between the current year and the previous year (**(dif)); and finally (d) a relationship based on the average change over the past three years (**(MA)).

The following hypotheses have been tested: With respect to the relationship with the economic business cycle the relationship of internal migration intensity is expected to be:

- positive with GDP per capita (GDP);
- negative with unemployment (UNEMP);
- positive or negative with inflation (house-owners profit from inflation as the real value of the loan on mortgage will decline: a positive relationship; or the other way round: although house-owners profit from inflation, increasing inflation may be a forerunner of a weakening economy: a negative relationship) (INFL);
- negative with interest rates on loans (lending interest, LEN);
- negative with real interest rates, i.e. interest rates on loans minus inflation (INTR); and
- positive with employment (EMP).

With respect to more structural economic developments, negative relationships are assumed with:

- the percentage of female labour force in the total labour force (LABF);
- the percentage employment in services in total employment (EMPS); and
- the ageing of the labour force, calculated as the percentage of the population aged 45-64 in the age group 20-64 (AGE).

[3] **: variable name; i.e. GDP(-1); UNEMP(-1) and EMP(-1)

Apart from possible economic influences, we expect an endogenous effect of internal migration itself (CMR(-1)), as internal migration behaviour at time *t* might be affected by internal migration at time *t*-1. This is due to the effect of vacancy chains on moving, where one change of residence opens up a new vacancy, which in turn triggers another move, and so on, until a starter in the housing market closes the chain (since a starter leaves no dwelling vacant).

Regression models

The following strategy has been used. First, we tested models with just CMR(-1) as the explanatory variable. Subsequently, we estimated models with CMR(-1) and one economic variable at a time. Economic variables were tested only if there was no collinearity with CMR(-1) (correlation with CMR(-1) below |0.8|). Due to multicollinearity as well as the limited number of observations, we could not test models including all economic indicators. Nevertheless, we tried to find out whether models with several economic indicators could better describe internal migration patterns compared to models including only one economic variable. Furthermore, we tested models *without* the endogenous internal migration effect to examine whether models with exclusively economic indicators could describe internal migration developments as well as, or even better than, models capturing the endogenous effect. Finally, for each country the stability of the best fitting model was tested using the Chow breakpoint and forecast tests, as well as the N-step forecast test.

From the analyses, we may conclude the following:

- The endogenous effect is strong for Italy, the Netherlands and Switzerland: R^2>0.90. Given the high correlations between CMR and CMR(-1), this is not surprising. For Spain too, a reasonable fit was found: R^2>0.80. For Finland and Sweden, the endogenous effect is less important, but still explains about half of the variance.
- Looking at the labour market indicators, we added one or more unemployment variables to the models of the Netherlands, Italy and Switzerland, and one or more GDP per capita variables for Finland and Sweden. For Spain we tried models with either unemployment, GDP per capita, or employment. Only for the Netherlands (UNEMP) and Spain (EMP(dif)), including one of the economic variables somewhat improved the basic model including only CMR(-1) (significant coefficient and higher R^2).
- As far as the other economic business cycle determinants are concerned, some effect was found for Sweden (inflation (negative) and lending interest), Finland (lending interest), Italy (inflation (positive)) and Spain (inflation (negative)). Adding real interest rates did not improve any of the models.
- Given the results of the univariate regression models and multicollinearity between the explanatory variables, combining two or more economic indicators was only possible for Finland and the Netherlands. For Finland a model combining lending interest and yearly changes in GDP per capita resulted in a sound description of internal migration intensities (significant coefficients for all variables; R^2=0.89). For the Netherlands, on the other hand, combining several indicators did not improve the model.

An overview of the best fitting models including CMR(-1) is given in Table 5.11. From this table we may conclude that no single model specification describes internal migration intensities over time across the given subset of European countries. The best model in terms of adjusted R^2, was found for Italy (R^2: 0.96). Although the best model for Sweden was less convincing than those for the other countries, still 70 per cent of the variation over time was explained by the model.

Table 5.11: Goodness of fit and coefficients of best fitting models including CMR(-1)

Country	Indicator	Coefficient	Standard error	t-value	Adjusted R^2
Finland	CMR(-1)	0.3298	0.136	2.43	0.89
	GDP(dif)	0.1076	0.037	2.88	
	LEN(-1)	-0.0402	0.015	-2.74	
Sweden	CMR(-1)	0.4780	0.170	2.82	0.70
	INFL(-1)	-0.0239	0.006	-3.78	
Netherlands	CMR(-1)	0.8128	0.047	17.15	0.93
	GDP(dif)	0.1262	0.044	2.85	
Switzerland	CMR(-1)	0.8709	0.066	13.29	0.92
Spain	CMR(-1)	0.8554	0.096	8.93	0.87
	GDP(dif)	0.1676	0.075	2.25	
Italy	CMR(-1)	0.7134	0.046	15.41	0.96
	INFL	0.0036	0.001	3.03	

(-1): a 1-year lagged effect

Table 5.12 shows the best fitting models excluding the endogenous effect of internal migration. Here we see that the models for Finland, Sweden and Spain without CMR(-1) are at least as good as models including the endogenous effect. For these countries, internal migration propensities seem to be affected mainly by the economic business cycle. More long-term structural effects, either endogenous or exogenous, seem to be limited. For Switzerland and Italy, on the other hand, we found a structural decline in internal migration which was hardly influenced by short-term economic developments. This decline might be related to the increase in the service sector (Italy), or the rising labour force participation of women (Switzerland). We have to note here, however, that this result was not surprising given the almost linear course of both internal migration intensity (declining) as well as female labour force participation in Switzerland (increasing) and employment in services in Italy. In fact, for those countries, all (economic) variables with a more or less linear course would have led to some kind of relationship. Interrelationships with ageing are less obvious.

Table 5.12: Goodness of fit and coefficients of best fitting models excluding CMR(-1)

Country	Indicator	Coefficient	Standard error	t-value	Adjusted R^2
Finland	LEN(-1) GDP(MA)	-0.0523 0.1717	0.009 0.044	-5.94 3.93	0.91
Sweden	INFL(-2)	-0.0416	0.003	-13.44	0.92
Netherlands	UNEMP(-1) LEN(-2)	-0.0216 -0.0171	0.005 0.003	-4.70 -6.07	0.79
Switzerland	LABF LEN(-1)	-0.1731 -0.0437	0.010 0.010	-16.81 -4.43	0.95
Spain	INFL(-2)	-0.0267	0.001	-20.10	0.92
Italy	EMPS	-0.0154	0.001	-21.29	0.96

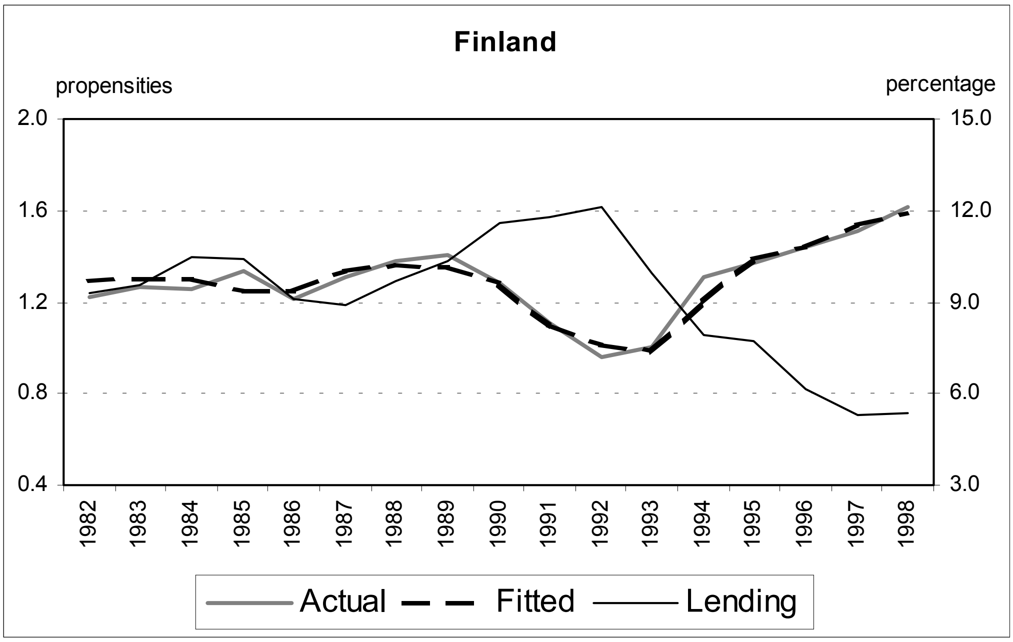
Results for the Netherlands are somewhat more complicated. For this country it is difficult to find a model in which the dynamic effect could be replaced by exogenous economic indicators. Although the outcome of a model with lagged unemployment and lending interest rates was quite reasonable, the overall fit was substantially lower than that of the best model capturing CMR(-1). However, as lending interest rates were only available from 1978 onwards, both models refer to different time periods. In the Netherlands, from 1972 to 1979 a strong decline of internal migration propensities took place. This period determines the strong endogenous effect. If we look at the limited period of 1980-1998, the best fitting model including CMR(-1) also captures Len(-2) and Unemp(-1). This model explains 83 per cent of the variance over time.

To sum up, we may state that for Italy and Switzerland, developments in internal migration propensities mainly follow a structural trend, either endogenous or exogenous driven. The same applies to the Netherlands in the period 1972-1979. In the other countries, as well as in the Netherlands from 1980 onwards, internal migration propensities seem to react in particular to fluctuations connected with the economic business cycle. As illustration, the observed and fitted migration patterns are given in Figure 5.14.

Stability of the models

To examine whether the parameters of the models are stable across various sub-samples of the data, we used the following tests: the Chow breakpoint test, the Chow forecast test and the N-step forecast test. According to the Chow breakpoint test the same models are estimated for different sub-samples of the data. The test statistics are based on a comparison of the sum of squared residuals obtained by estimating the model to the entire sample, with the sum of squared residuals obtained when the models are fit to each of the sub-samples. Significant differences between the models indicate a structural change of the relationship. The Chow forecast test, on the other hand, estimates the model for a subset of the data and uses the estimated relationship to predict the values of the dependent variable in the remaining period. Large differences between predicted and observed values of the dependent variable indicate a changing relationship. Both tests do not necessarily have the same results: models may be stable, but not satisfactory as far as prediction concerns, or models may change over time, but may result in stable predictions.

Figure 5.14a: Actual and fitted internal migration propensities (%), Finland and Sweden



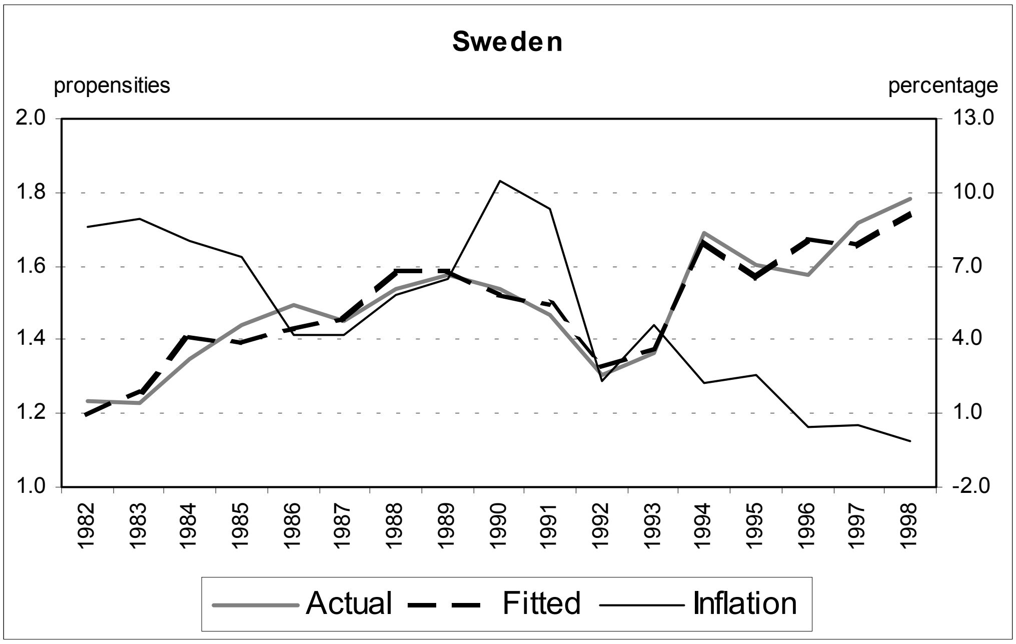
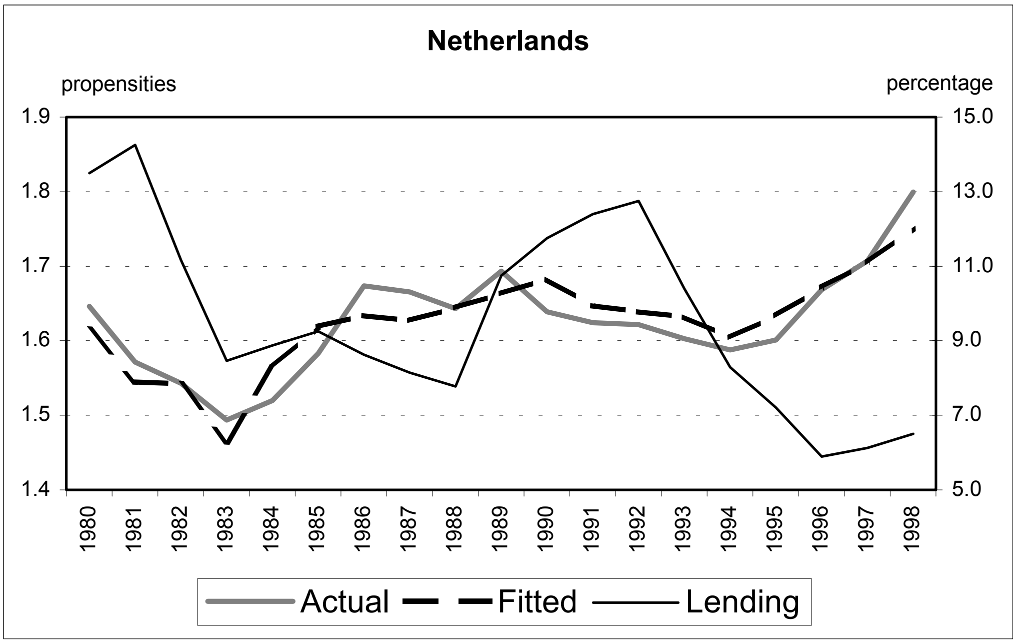


Figure 5.14b: Actual and fitted internal migration propensities (%), the Netherlands and Switzerland



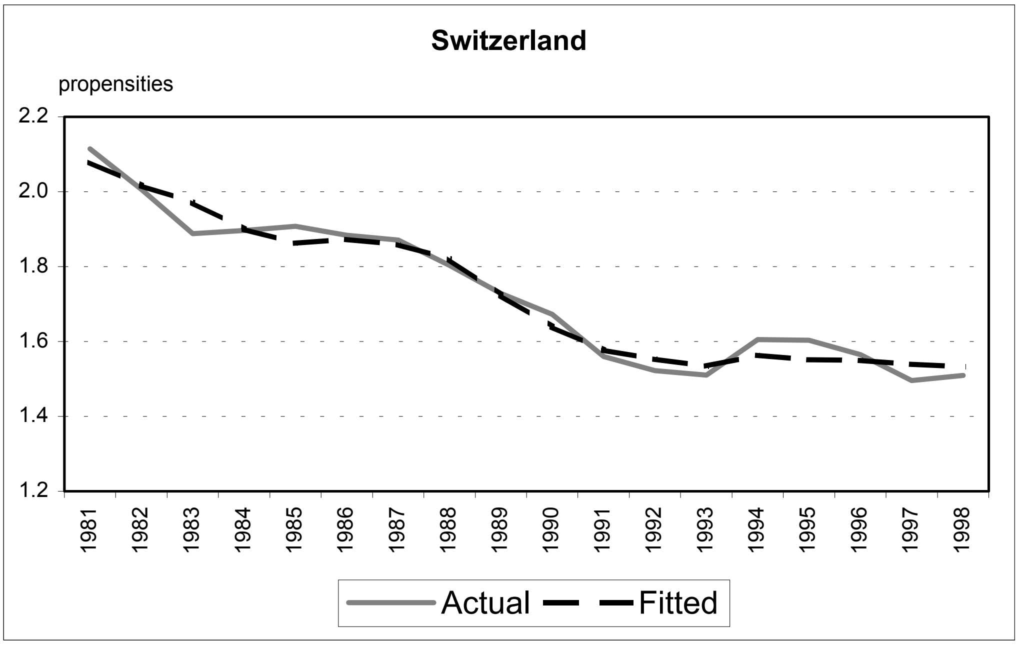
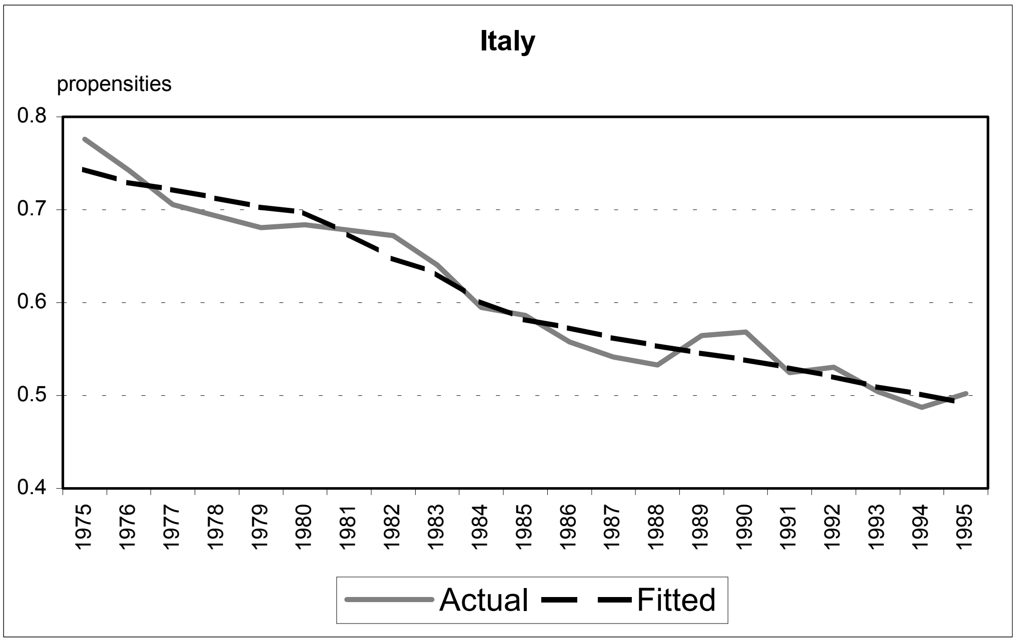
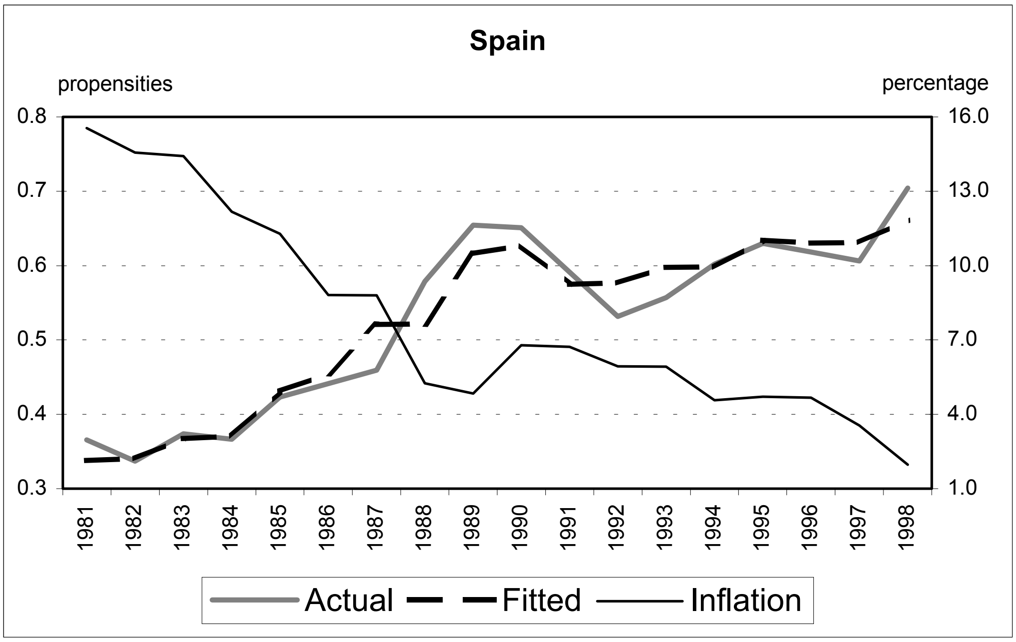


Figure 5.14c: Actual and fitted internal migration propensities (%), Italy and Spain





Following the N-step forecast test, a sequence of Chow forecast tests is carried out. Instead of specifying one single observation and forecast period, all possible sample sizes are tested, starting with the smallest observation subset possible (number of observation equal to number of coefficients in the model) predicting all other values, and ending with a subset of all-but-one observations to predict the final, most recent, value. The result of this test is a plot of recursive residuals about the zero line. Each recursive residual is the error in a one-step ahead forecast test. Residuals outside the two standard error bands suggest instability of the (one-step ahead) model. In addition, the plot shows the significant N-step probabilities for those points where the hypothesis of stability would be rejected. For the Chow Breakpoint test we used two different breakpoints. First, we divided the total sample in two periods of equal lengths. Second, we split the period in a long-term 'base' period, followed by a short period of four to five years (breakpoint: 1995, for Italy: 1990), to test whether the relationship has recently changed. For all countries, we examined the stability of both models, model 1 including the endogenous migration term (CMR(-1)), and model 2 excluding CMR(-1).

The results indicate the following:

- According to the Chow Breakpoint test, for Spain a significant result was found for model 1, breakpoint 50%-50%. For the Netherlands (breakpoint 50%-50%), and Switzerland (breakpoint 1995) a significant result was found for model 2. In all other cases, parameter estimations were quite stable across different sub-samples of the dataset.
- Looking at the recursive residuals, data points outside the boundaries of two standard errors are limited. Only for Sweden (model 1, 1994), Switzerland (model 2, 1997) and Spain (model 1, 1993; model 2, 1988) one or two years were not predicted satisfactory.
- Quite some more sample points have N-step probability values at or below 15 percent. For both models for Sweden there is evidence of instability in the first half of the 1990s. The same applies for Spain in the second half of the 1980s and for model 2 for Switzerland for the period 1993-1996. For the other countries, at most one of the p-values is below 0.15. As this value is still well above 0.5, the estimated relationships for the other countries are rather stable over time.

Summarizing we may conclude that the results of the models with cyclical indicators were fairly consistent, while the models with structural indicators were rather mixed. This is particularly striking, as in theory the structural indicators included in this study (labour force participation of women, the rise of the service sector and the ageing of the labour force), seem to be especially important with respect to long-distance migration, while the cyclical indicators, for instance GDP per capita, unemployment, and employment, may be important for both short-distance as well as long-distance migration. Nevertheless, some support was found for all of the hypotheses tested. Furthermore, for all countries one or more good fitting models were found. In general, these models were rather stable over time and predictions based on these models were mostly satisfactory. On the other hand, we did not find robust relationships across a number of countries. A possible reason for this is that internal migration is strongly related to the institutional context of the housing market in the different countries (financing rules, tax systems, building construction). Moreover, developments on the housing market may also obscure the relationship between internal migration propensities and economic indicators. For instance, in times of economic growth, rents and housing prices may grow rapidly, which may put constraints on the level of migration.

5.5 General conclusions about modelling internal migration

In this section we sum up the general conclusions of the modelling exercise:

- The results of the demographic and economic models are broadly similar for each country.
- For out-migration we found broadly similar specifications of the economic models. Unemployment and GDP per capita are significant explanatory variables for out-migration, although their explanatory value is limited. For destination choice, GDP per capita was found to be important to model the attractiveness functions of regions.
- Despite these broad similarities the coefficients of the variables show a large variation over countries. For the Netherlands for out-migration we needed another definition of the variables GDP and unemployment (regional differences compared to the national values) and a comparison of coefficient values reveals that there is no robust specification that fits all countries. One robust model could not be found.
- There are large differences in migration behaviour between age groups, as reflected in the model parameters.
- The added explanatory value of economic variables is limited. Internal migration processes are relatively stable in terms of age and sex patterns, and in terms of regional differences and interaction flows.
- Despite this limited explanatory power of economic variables for internal migration, in Sweden and the Netherlands models using economic variables and a distance function were better predictors of destination choice probabilities than purely demographic models.
- There is a systematic under-prediction in almost all out-migration models. The explanatory models using economic variables do not reduce the bias. In other words, perfect knowledge about future GDP, regional unemployment and population density do not improve the forecast.
- Change over time in internal migration intensities is only to a limited extent explained by economic variables. In some countries internal migration mainly follows a structural trend, either endogenous or exogenous driven. In other countries internal migration propensities seem to react in particular to fluctuations connected with the economic business cycle.
- In general, models of migration intensities were rather stable over time and predictions based on these models were mostly satisfactory. However, we did not find robust relationships across a number of countries. For consistency reasons, therefore, (short term) predictions of internal migration intensities are best performed by imposing a time trend, either using time-series results or based on exogenous information.
- The overall conclusion is that internal migration is a complex phenomenon with many national and regional idiosyncrasies. Therefore a "one size fits all" method is not feasible for all countries. It does seem to be feasible, however, to formulate a general framework for scenarios to be used in several countries, but a country-specific implementation will be needed to take into account all country-specific features of relevance. This means that we move from a "one size fits all *model*" to a "one size fits all *approach*" with country-specific interpretation.

6 EU-wide implementation

In the previous chapter, migration has been modelled as a two-stage process with first the out-migration from origins and second the distribution of migrants between destinations. In both stages economic variables were explicitly used to explain and predict interregional migration. The subject of the current chapter is to show how the results of the analyses could be used in formulating a general framework of sub-national population projections for the countries of the European Union. Section 6.1 gives the outline of the calculation scheme of the two-stage model proposed in the current study, further indicated as the REGMIG model (REGional MIGration).

In order to compile demographic scenarios based on socio-economic indicators assumptions are needed on how the socio-economic indicators might change in the future. Here several approaches can be followed. We can either use linear trend analyses of the explanatory variables, link the demographic scenarios to (existing) economic scenarios, or formulate so-called 'what-if' scenarios, in which we can project demographic consequences of hypothesised economic developments, such as a growth in regional incomes by one percent. The process of scenario-making is subject of section 6.2.

A practical issue in setting up scenarios is whether we should formulate assumptions for all regions separately, or whether it is more efficient to develop scenarios for only a limited number of groups of regions. In the latter option, we have to group regions together with more or less similar characteristics on the socio-economic variables. In section 6.3 we investigate whether it is possible to construct a meaningful classification of NUTS 2 regions, which can be used both at the European level as well as within individual countries. In section 6.4 we will have a closer look at regional disparities in economic developments.

So far, data have been analysed for four case-study countries only: Sweden, the Netherlands, Spain and the United Kingdom. In order to develop a European wide model of interregional migration, we need to know to what extent the results of the analyses could be generalised to the other EU countries. A major constraint in this respect is data availability. Therefore, in section 6.5 an overview will be given for the current EU countries of all NUTS 2 level internal migration data available in Eurostat's database, New Cronos. Although the current study was planned to make use of New Cronos, it was decided to obtain demographic data from the national statistical offices as age-specific data on origin-destination flows were not available in New Cronos. Obviously, this will be the same for the other EU-countries, therefore, additional information has been included on data availability at the NSOs (see also chapter 3). Finally, some remarks will be made on the availability of the explanatory variables.

In section 6.6, finally, a summarising overview will be given of all steps to be taken to produce internal migration scenarios using the REGMIG model.

6.1 Calculation scheme

In the EUROPOP1995 model, a limited number of scenario parameters were defined, based on the multidimensional structure of internal migration. With the exceptions of the total mobility level and the aggregate origin and destination effects, all parameters were held constant. Consequently, the total mobility level affected all internal migration rates in the same way and the aggregate origin/destination effects influenced all migration flows, irrespective of age and sex. In this way, future rates could be calculated as:

$$r_{ijast} = t \cdot o_i \cdot d_j \cdot v_{ias} \cdot w_{jas} \cdot m_{ij} \qquad (6.1)$$

in which:
r_{ijast} = internal migration rates by age/sex/origin/destination;
t = multiplication factor to reach the hypothesised future mobility level;
o_i = aggregate origin effects;
d_j = aggregate destination effects;
v_{ias} = out-migration rates by age and sex;
w_{jas} = destination shares by age and sex; and
m_{ij} = aggregated OD-effects

The REGMIG framework can be considered as an upgrade of the EUROPOP1995 model in the sense of explicitly including socio-economic indicators and age-specific origin-destination patterns. Separate models have to be used for out-migration and destination choice. In the REGMIG framework future rates could be calculated as

$$r_{ijast} = o_{iast} \cdot p_{j|asti} \qquad (6.2)$$

in which
o_{iast} = age and sex specific origin effect; and
$p_{j|asti}$ = age and sex specific destination effect, conditional on region specific out-migration

Equation (6.2) contains two different parts representing the origin and destination choice effect. Comparing both models we see that parameters o_i and v_{ias} of formula (6.1) are captured in equation (6.2) in the origin effect o_{iast} (or out-migration model), while the parameters d_j, w_{jas} and m_{ij} are captured by the destination effect $p_{j|asti}$.

The origin effect o_{iast} is composed of three components: 1) a region specific age and sex profile which is not dependent on time, 2) an overall level of migration which is following an autonomous trend, and 3) an economic component in which the relationship with socio-economic variables is modelled.

The origin effect o_{iast} could be calculated as

$$\begin{aligned} o_{iast} &= \exp(\gamma_{ias} + \alpha_t + X_{iast}\beta_{as}) \\ &= \Gamma_{ias} \cdot A_t \cdot \exp(X_{iast}\beta_{as}) \end{aligned} \qquad (6.3)$$

in which

Γ_{as} = region-specific age/sex profile;
A_t = overall base rate (migration level);
X_{iast} = a vector of explanatory variables, particular to region *i,* age/sex *a,s*, and time *t*; and
β_{as} = a vector of coefficients, which may be age- and sex-(*a,s*) specific.

In equation (6.3) β_{as} could also be expressed in terms of elasticities , $\in_{as}$: the percentage change in o_{iast} due to one percentage change in a given *X* variable.

Although not captured in equation (6.3), the overall migration level might be related to economic indicators as well. This has been modelled separately. Apart from economic influences, internal migration intensities at time t could be affected by the endogenous effect of internal migration itself (intensities at time *t*-1). This endogenous effect of internal migration has also been taken into account.

The overall migration level could be calculated as

$$A_t = \lambda A_{t-1} + X_t\beta \qquad (6.4)$$

In which

A_{t-1} = the endogenous effect of internal migration;
X_t = a vector of national level socio-economic indicators; and
β = a vector of coefficients.

The destination choice effect, finally, can be decomposed into an economic and a spatial component. This destination effect $p_{j|asti}$ was modelled as a multinomial logit model, with the following specification:

$$p_{j|iast} = \frac{\exp(\alpha_{ias} + X_{jt}\beta_{as} + F_{ijas})}{\sum_{k=1}^{J}\exp(\alpha_{kas} + X_{kt}\beta_{as} + F_{ikas})} \qquad (6.5)$$

where

α_{ias} = a destination-specific constant, capturing all non-variant attractiveness elements in destination *j* which are not taken into account in the explanatory variables $X_{it;}$
X_{jt} = a vector of explanatory variables for destination *j* at time *t*;
β_{as} = a vector of coefficients, which may be age/sex-specific; and
F_{ijas} = an interaction term between origin *i* and destination *j*, which may be age/sex-specific.

The interaction term F_{ijas} may be in the form of a historical migration matrix OD_{ij}, or in the form of a distance function. From analyses, it became clear that a historical migration matrix gives better predictions.

6.2 Scenario-making

In general, for scenario purposes we can distinguish between a purely demographic model, a purely economic model and a model, which is driven by both demographic and economic developments, a so-called mixed model. Given the results of the current study, we propose to use a mixed model, with socio-economic indicators for the region specific component of out-migration and the regional attractiveness. For the spatial component of the destination choice model we suggest to use historical migration patterns (the origin-destination (OD) matrix and for the overall level of migration we propose to impose a time trend. For the four case study countries the relevant explanatory variables for each of the different parts in the framework are summarised in Table 6.1.

Table 6.1: Country-specific explanatory variables

Country	Out-migration		Destination choice	
	Overall level of migration	Region specific component	Spatial component	Regional attractiveness
The Netherlands	Time trend	LPOP GDPZLAG, UNEMPZLAG A.DENS	OD	A.GDPLAG
Sweden	Time trend	LPOP A.GDPLAG A.UNEMPLAG A.DENS	OD	A.GDPLAG
The United Kingdom	Time trend	LPOP A.GDPLAG A.UNEMPLAG A.DENS	OD	A.GDPLAG
Spain	Time trend	LPOP A.GDPLAG A.DENS	OD	

From Table 6.1 we may conclude that for the out-migration and destination choice models regional population, GDP per capita, unemployment and population density are important socio-economic indicators. In all cases, country-specific specifications are needed.

In contrast to the EUROPOP1995 model, in which hypotheses were formulated on future developments of the demographic parameters, in REGMIG hypotheses have to be formulated on future developments in the economic variables. Here, several options are possible:
Use trend extrapolations of the X-variables (for instance GDP per capita, or unemployment rate).
Link the population scenarios to (existing) economic scenarios. In this latter case we could, for instance, fit in with European scenarios developed by the Central Planning Office (CPB) of the

Netherlands. De Mooij and Tang (2003), and Lejour (2003) present four European scenarios until 2040 (see below).
Develop high and low bounds of migration based on observed extreme values observed in the (recent) past.

As an example of the second scenario option, the CPB office recently developed a set of long term economic scenarios. They distinguish two basic dimensions or key uncertainties: the first challenge is whether countries will succeed in international cooperation, while the second refers to institutional reforms in the public sector. Those two dimensions result in four basic scenarios, called Strong Europe, Global Economy, Regional Communities and Transatlantic Markets. In Strong Europe and Global Economy international cooperation is a prominent theme, while in the other two scenarios international cooperation is only limited. The role of the public sector is most prominent in Strong Europe and Regional Communities. Assumptions on GDP per capita and unemployment developments in the four scenarios for the EU15 are given in Figure 6.1.

Figure 6.1: Four scenarios for Europe (CPB)

International cooperation			
Public responsibilities	**Strong Europe** Third growth in GDP Moderate unemployment	**Global Economy** Highest growth in GDP Low unemployment	Private responsibilities
	Regional Communities Lowest growth in GDP High unemployment	**Transatlantic Market** Second growth in GDP Low unemployment	
National sovereignty			

Source: Lejour (2003)

Although European economic scenarios such as the CPB scenarios could be useful as context for the REGMIG scenarios, most often, economic scenarios at the European level will not provide detailed input for each of the countries separately, not to mention input at the regional level. Therefore, additional assumptions on country and region-specific implementation will be necessary. Alternatively, instead of using European-wide scenarios, one can try to adopt country-specific economic scenarios. It remains to be seen however, to what extent country-specific scenarios will be based on a set of consistent assumptions.

In producing so-called 'what-if scenarios', there are fewer restrictions on the assumptions on future economic developments. In this case we could take into account the consequences of hypothetical changes in economic variables. What-if scenarios could be used for instance to answer questions such as: what will happen to migration and population if regional incomes will grow by one percent, or if you double income inequalities? Nevertheless, for the final set of scenarios to be used in this project choices have to be made about these values.

Whether we adopt existing economic scenarios or formulate what-if scenarios, in both cases we should make a decision on whether we should formulate assumptions for all regions separately, or whether it will be more efficient to develop scenarios for a limited number of groups of regions only. In the following section, we will pay attention to the possibilities and impossibilities of classifying groups of individual NUTS 2 regions.

6.3 A classification of NUTS 2 regions

One of the conclusions of chapter 5 is that GDP per capita and unemployment are important determinants of interregional migration. In addition to these economic determinants, population density also affects the direction of migration flows. In order to simplify the specification of time trends in these key indicators for demographic scenarios, we used cluster analysis to produce groups of similar regions.

We have carried out a cluster analysis for all EU15 countries at NUTS 2 level using the variables population density, GDP per capita and unemployment rate converted into rank values to avoid the problem of concentration of a majority of units into a big central cluster. This analysis resulted in four clusters. Cluster 1 and 2 are characterised by, on average, low incomes, high unemployment and low or high population density; while clusters 3 and 4 are characterised by high income, low unemployment and high or low density respectively (see table 6.2).

Table 6.2: Average economic indicators for four clusters of regions (based on K-means cluster analysis)

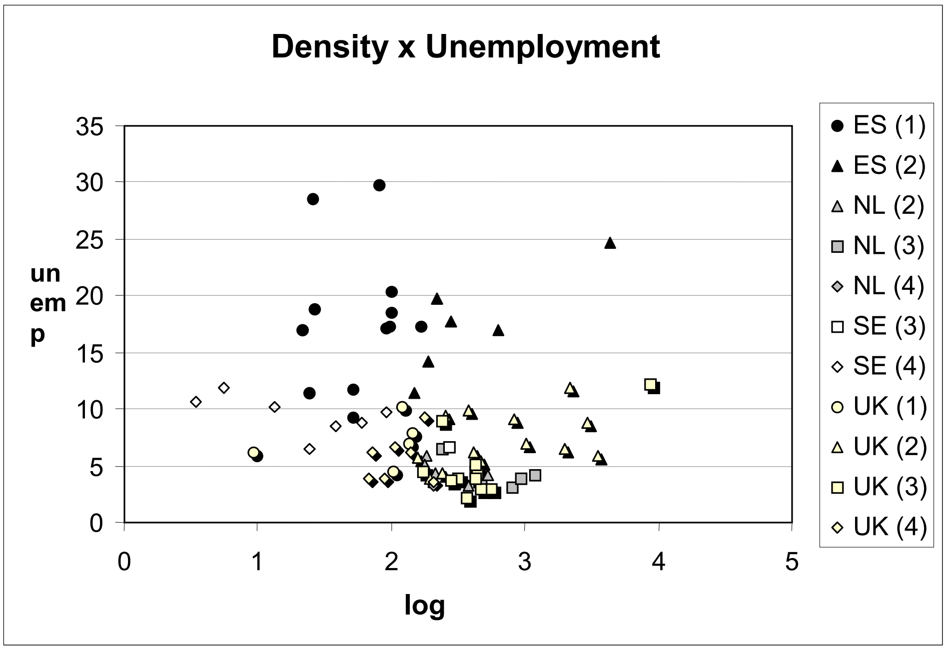
	Number of regions	Average population (x 1000)	GDP	GDP per capita	Unemployment	Density
cluster						
1	61	1223.91	16255.02	13.34	13.30	74.18
2	46	1924.54	30992.02	16.29	12.03	628.73
3	52	2489.27	65847.31	26.23	6.65	855.56
4	52	1611.51	34584.18	21.72	7.58	98.80

If we look at the classifications of NUTS 2 regions for the four case-study countries only (see Figure 6.2 in which unemployment and GDP per capita are plotted against population density), we see a clear difference between the two low density clusters (clusters 1 and 4, indicated by circles and lozenges) and the two high density clusters (clusters 2 and 3, indicated by triangles and squares). For the variables GDP per capita and unemployment, the pattern is less clear. Although on average for all EU15 countries there seems to be a distinction between high and low GDP per capita, for these four countries all clusters seem to contain regions with relatively high and low GDP. For unemployment, this is the case for clusters with on average high unemployment (1 and 2), but not for the low unemployment ones. Looking at the differences between countries, we see a clear country effect: all high unemployment regions belong to Spain (black symbols), while for Sweden (blank symbols) all but one region fall in the low density/low unemployment cluster and for the Netherlands (grey symbols) almost all regions belong to one of the high density clusters. The UK (shadowed symbols) is the only country with regions in all four clusters. However, the results for the UK do not match another NUTS 2 ranked analysis carried out on the more detailed level of UK local authorities. Given these results, we are not convinced that this classification of regions creates meaningful groups of regions with similar socio-economic characteristics which can be used in the individual countries.

Alternatively, we classified the regions of the four case-study countries directly according to their values on the dimensions of population density and GDP per capita (Figure 6.3). Here we used the percentage difference compared to the national value, averaged over the total period 1990-1998. For density, this is not much different from the value of the most recent year; for GDP per capita, however, there are some minor differences but in general both patterns are highly similar.

Looking at Figure 6.3 we may conclude that a country-specific classification of regions in four quadrants of high/low density and high/low GDP per capita is also difficult to make. Although some similarities between regions in the same quadrant were found, for instance most of the main urban centres were classified in the high density - high GDP quadrant, differences between regions in the same quadrant may be larger than differences between regions in two different ones. In the United Kingdom, for instance, it doesn't seem to make sense to classify London together with the other regions in the high density - high GDP per capita group. In Sweden, the Stockholm area is so dominant that differences between the other regions are relatively minor and in Spain Islas Baleares (the Balearic Islands) are grouped together with Madrid, Cataluña and Pais Vasco, which seem to be quite different regions. Another difficulty in this respect is the definition of GDP. GDP per capita measures economic performance in terms of income generated in a region. This stresses the production side and not the user side of regional incomes. Production based and disposable incomes can vary substantially. In the Netherlands, for instance, Groningen shows a high GDP per capita but a relatively low per capita disposable income, while for Flevoland on the other hand the reverse is true. There are several reasons for this difference, among which commuting and transfers (income generated in one region may be allocated in another region) play an important part (Vanhove, 1999). So whilst GDP per capita seems to be a good indicator to be used in scenarios of economic prosperity, it seems to be more difficult to be used as indicator to classify NUTS 2 regions. Also a classification based on density and unemployment is difficult to understand (Figure 6.4). According to this classification, for instance, Noord-Holland, one of the Randstad regions of the Netherlands, is classified together with Limburg and not with the other Randstad regions Zuid-Holland and Utrecht.

Figure 6.2: Classification of regions using K-means cluster analysis



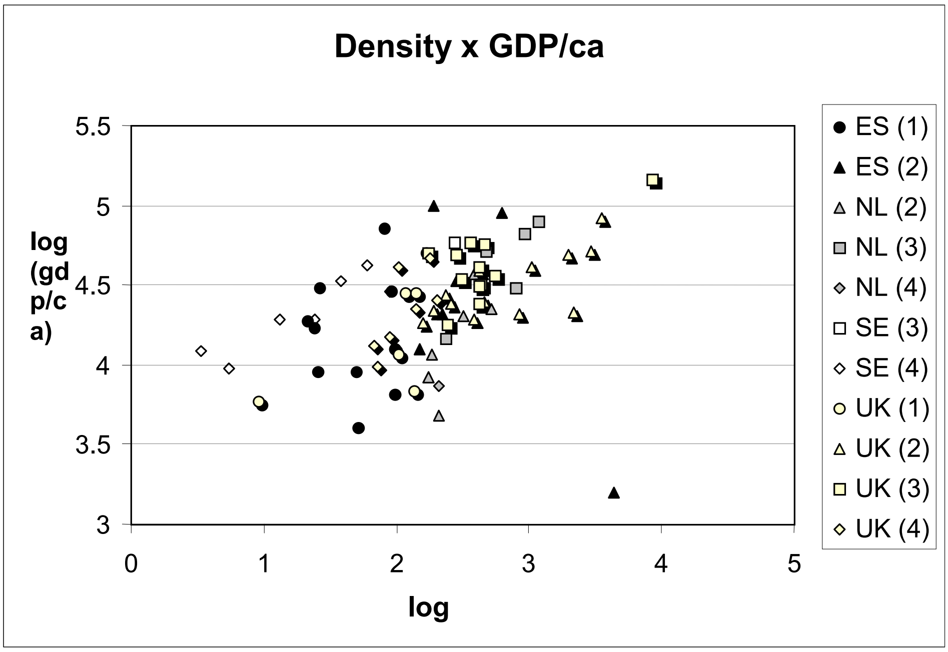
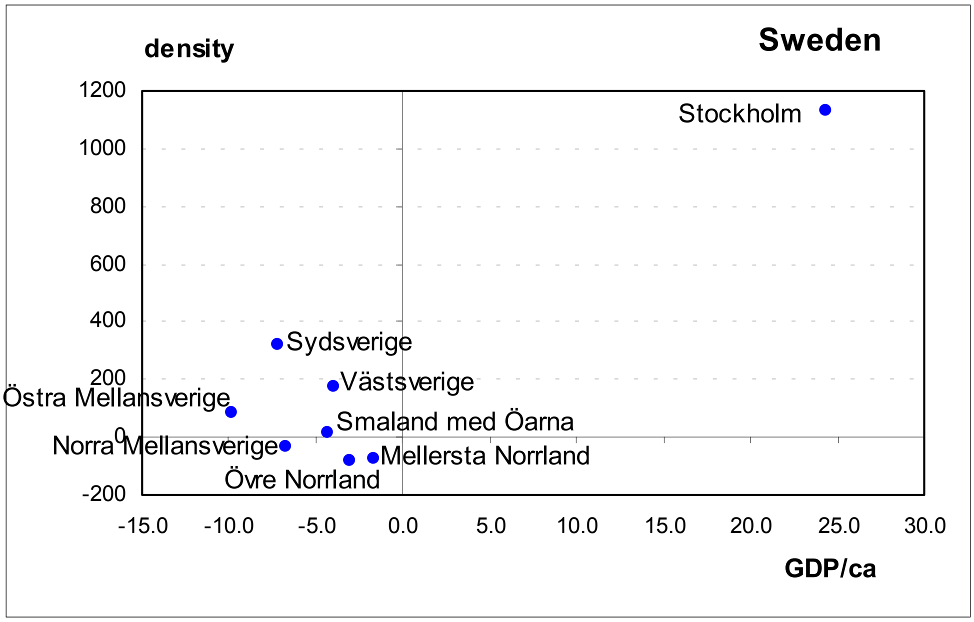


Figure 6.3: *NUTS 2 regions, GDP per capita and density (percentage difference to national level, 1990-1998), Sweden and Netherlands*



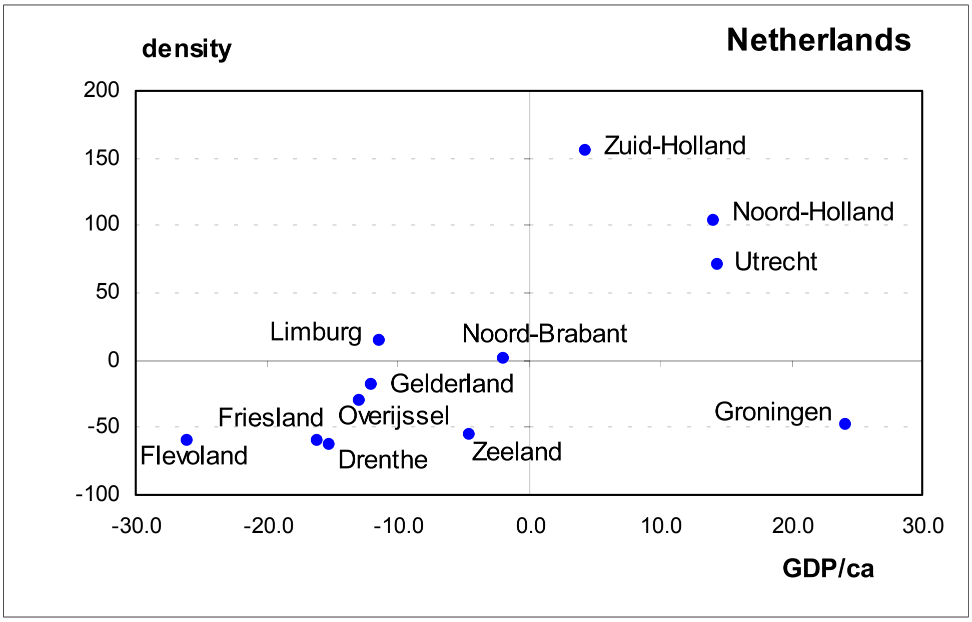
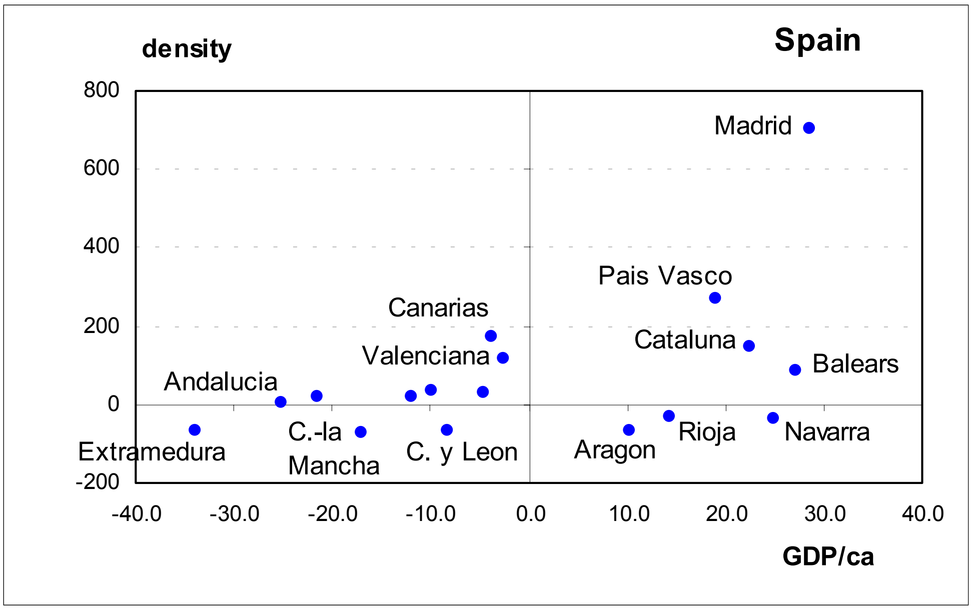


Figure 6.3: (continued) NUTS 2 regions, GDP per capita and density (percentage difference to national level, 1990-1998), Spain and United Kingdom



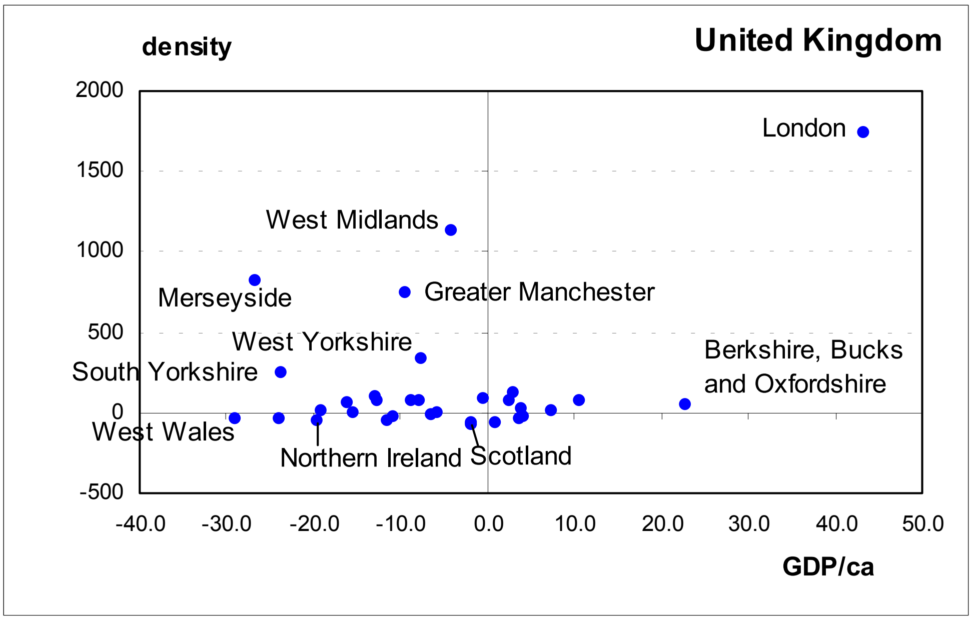
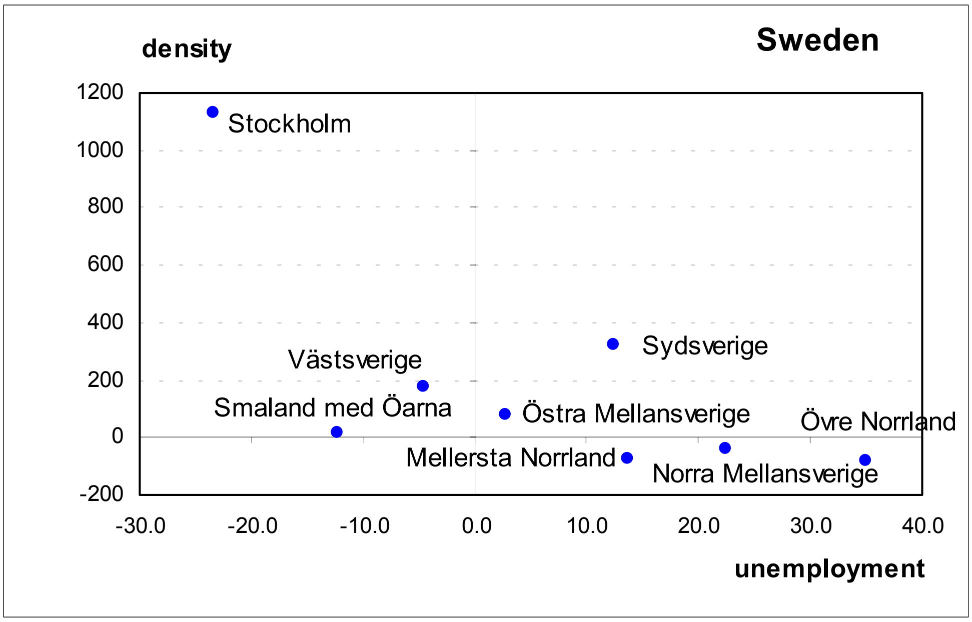


Figure 6.4: NUTS 2 regions, unemployment and density (percentage difference to national level, 1990-1998), Sweden and Netherlands



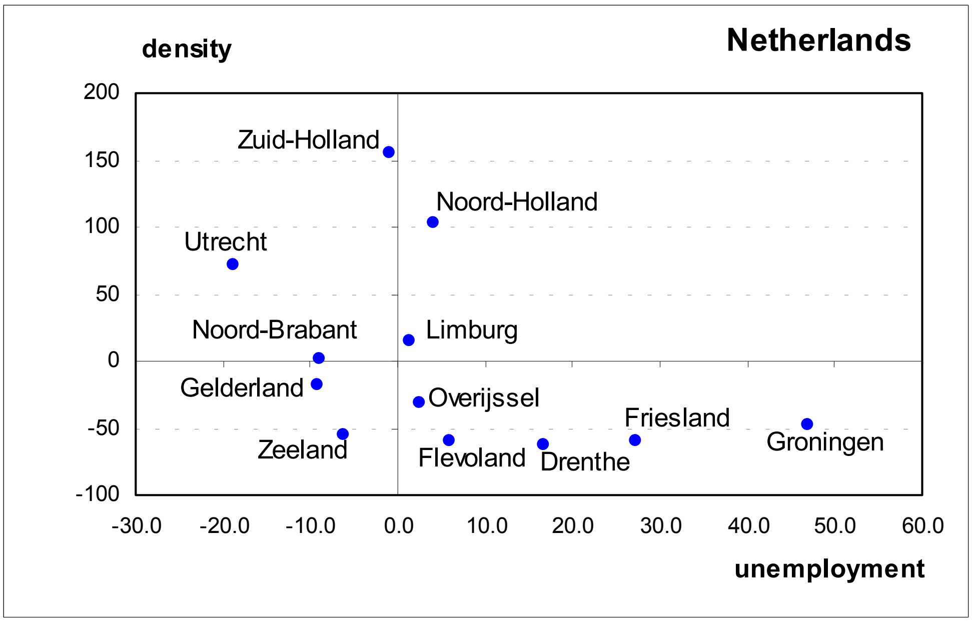
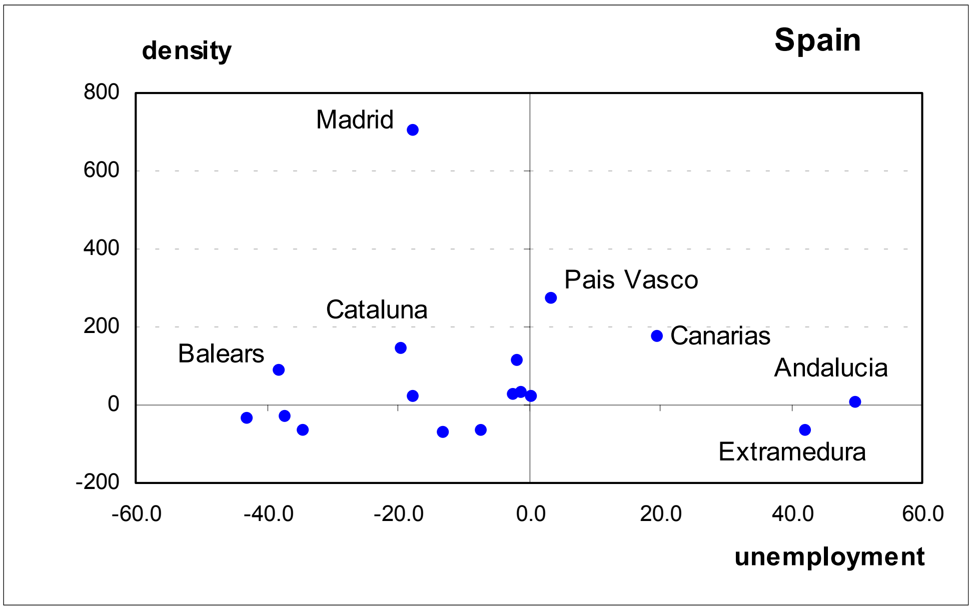
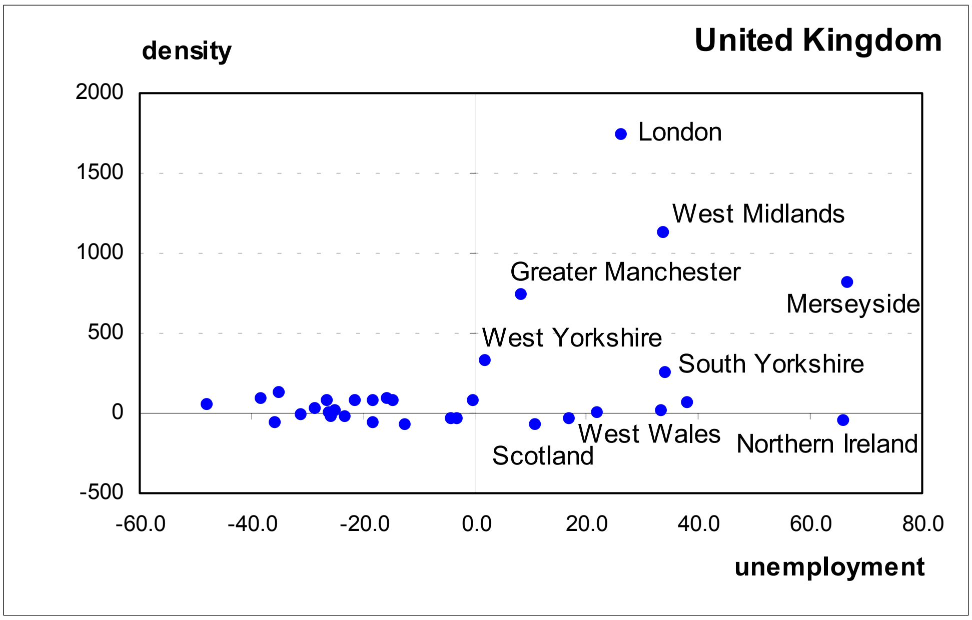


Figure 6.4: (continued) NUTS 2 regions, unemployment and density (percentage difference to national level, 1990-1998), Spain and United Kingdom





Given these results we came to the conclusion that classifying NUTS 2 regions on European scale is a difficult task to perform. For the given economic indicators and the geographical NUTS 2 level differences between countries are too large to find a meaningful classification of regions at the level of individual countries, which is needed for a country-specific implementation of the general model. Unfortunately, we cannot use cluster analysis at NUTS 2 level for the individual countries as for the majority of the countries the number of NUTS 2 regions is too small. And even if we have a large number of regions, as in the UK, the results of the analysis might be difficult to understand because of the huge variety in NUTS 2 regions within the country. Therefore, more detailed country-specific information is needed to come to a meaningful classification of NUTS 2 regions.

6.4 Regional disparities in economic development

In this section we look at the regional disparities in GDP per capita and unemployment within the four case study countries. In Table 6.3 national values of GDP per capita and unemployment are given together with the extreme values for NUTS 2 regions for Sweden, the Netherlands, Spain and the United Kingdom for the years 1990 and 1998. A first conclusion we may derive from this table is the difference between the countries in both GDP per capita as well as unemployment. Especially the economic position of Spain is lagging behind. GDP per capita is far below the level of the other three countries while unemployment levels are much higher. Differences in GDP per capita were somewhat larger in 1998 compared to 1990, while for unemployment we saw the reverse.

Looking at regional disparities within countries, the largest differences are found in the United Kingdom and Spain. This is not surprising as these countries have the largest number of NUTS 2 regions, and the higher the number of regions, the greater the disparities. For all countries disparities are especially large for unemployment. In 1990 in the United Kingdom, the ratio of unemployment in the region with the highest level (Northern Ireland) to the region with the lowest level (Berkshire, Bucks and Oxfordshire) is even more than 7. GDP per capita is about twice as high for the richest regions of the United Kingdom and Spain compared to the poorest regions. For the Netherlands and Sweden this ratio is about 1.5.

Table 6.3 takes into consideration only the extreme values within each country. In order to compare countries and periods more properly, we calculated the weighted coefficient of variation:

$$CV = \frac{1}{\bar{y}}\sqrt{\sum (y_i - \bar{y})^2 p_i / n} \qquad (6.6)$$

in which

CV = weighted coefficient of variation;
y = national level of GDP per capita or unemployment;
y_i = GDP per capita or unemployment in region i; and

p_i/n = population share of region *i*.

The weighted coefficient of variation *CV* is a measure of the dispersion of the observations as a whole (Vanhove, 1999). In Figures 6.5 and 6.6, disparities are plotted for GDP per capita and unemployment, respectively.

Table 6.3: Value ranges of GDP per capita and unemployment for NUTS 2 regions in Sweden, the Netherlands, Spain and the United Kingdom, 1990 and 1998

Country		Minimum		Maximum		National value
		Region	Index	Region	Index	
GDP		(country level = 100)				
1990	SE	Östra Mellansverige	91	Stockholm	120	15843
	NL	Flevoland	76	Groningen	122	14775
	ES	Extramadura	65	Balears	129	11271
	UK	West Wales and The Valleys	71	London	142	14404
1998	SE	Östra Mellansverige	89	Stockholm	131	21327
	NL	Flevoland	72	Utrecht	124	23384
	ES	Extramadura	63	Madrid	134	16068
	UK	Merseyside	70	London	148	20968
Unemployment		(country level = 100)				
1990	SE	Stockholm	55	Övre Norrland	197	1.6
	NL	Zeeland	71	Groningen	168	7.3
	ES	Rioja	45	Ceuta y Melilla	181	16.4
	UK	Berkshire, Bucks and Oxfordshire	34	Northern Ireland	247	7.0
1998	SE	Stockholm	76	Övre Norrland	124	6.5
	NL	Utrecht	79	Groningen	167	3.9
	ES	Madrid	49	Andalucia	157	19.0
	UK	Berkshire, Bucks and Oxfordshire	35	Merseyside	190	6.2

Source: Eurostat/ NSOs

Here we see that for all four countries income disparities in the 1990s slightly increased. For unemployment in the beginning of the 1990s decreasing disparities were observed, followed by an upward trend. Except for the Netherlands, the dispersion in unemployment was larger in 1990 compared to 1998. If we combine trends in disparities with the national levels of GDP per capita and unemployment (Figures 6.7 and 6.8) we find a positive relationship for GDP per capita and a negative one for unemployment (see Table 6.4). Thus, increasing levels of GDP per capita coincide with relatively higher regional disparities. Does this mean that with growing levels of GDP per capita regions lagging behind will profit least of economic improvements? Not necessarily. For the Netherlands, the United Kingdom and Spain, correlations between absolute levels of GDP per capita in 1990 and growth in 1998 were all below 0.3. If we look at the Netherlands, for instance, we see that both Flevoland and Groningen, the regions with the lowest and highest GDP per capita in 1990, experienced a growth in GDP per capita of 50 per cent. In Utrecht, however, in the period

1990-1998 GDP per capita grew with more than 80 per cent while the smallest increase was observed in Zeeland (a growth of 34 per cent). In Sweden, on the other hand, developments in Stockholm were absolutely dominant: GDP per capita in the capital region grew with 46 per cent while in the remaining regions grow percentages varied between 26 and 35 per cent, with 31 per cent for Östra Mellansverige.

Figure 6.5: Regional dispersion in GDP per capita, 1990-1998

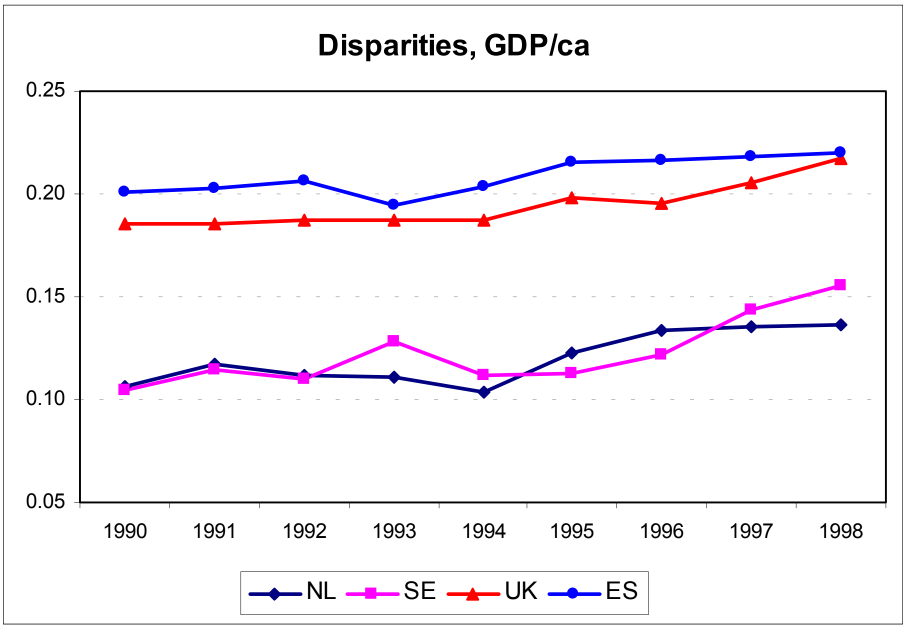


Figure 6.6 Regional dispersion in unemployment, 1990-1998

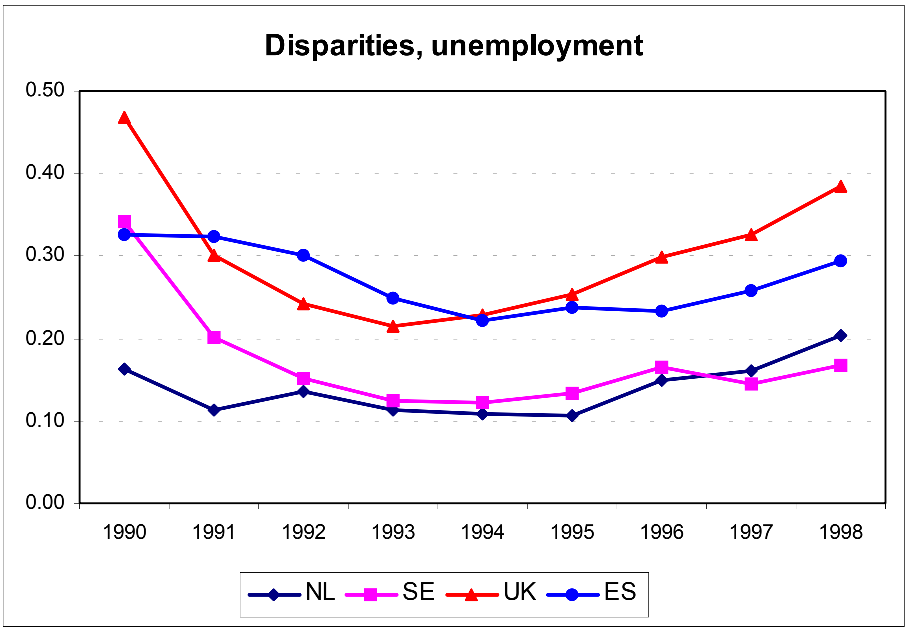


Figure 6.7: National levels of GDP per capita, 1990-1998

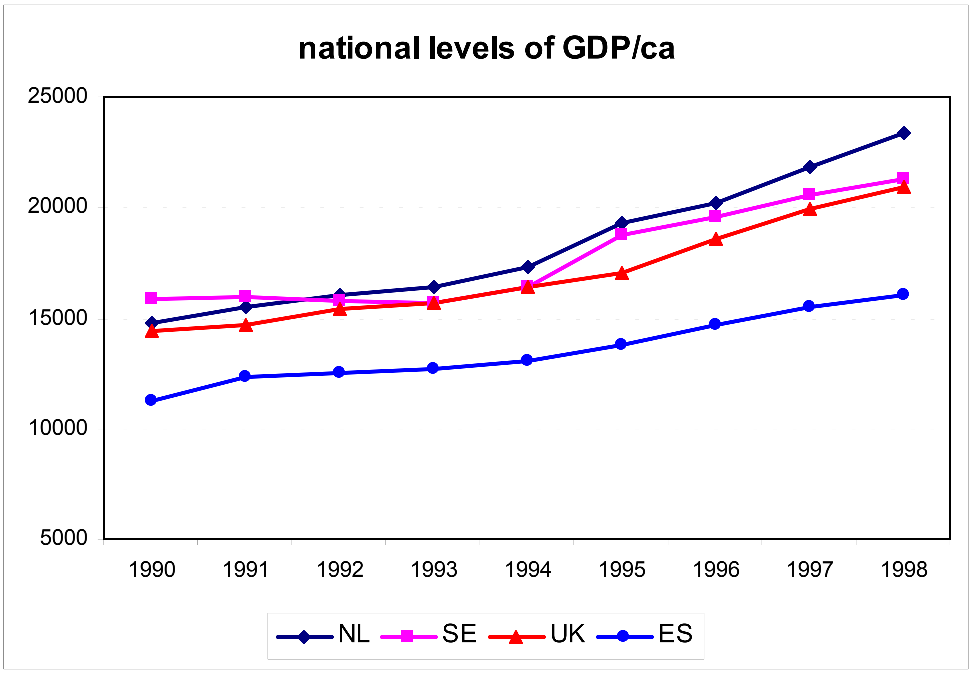


Figure 6.8: National levels of unemployment, 1990-1998

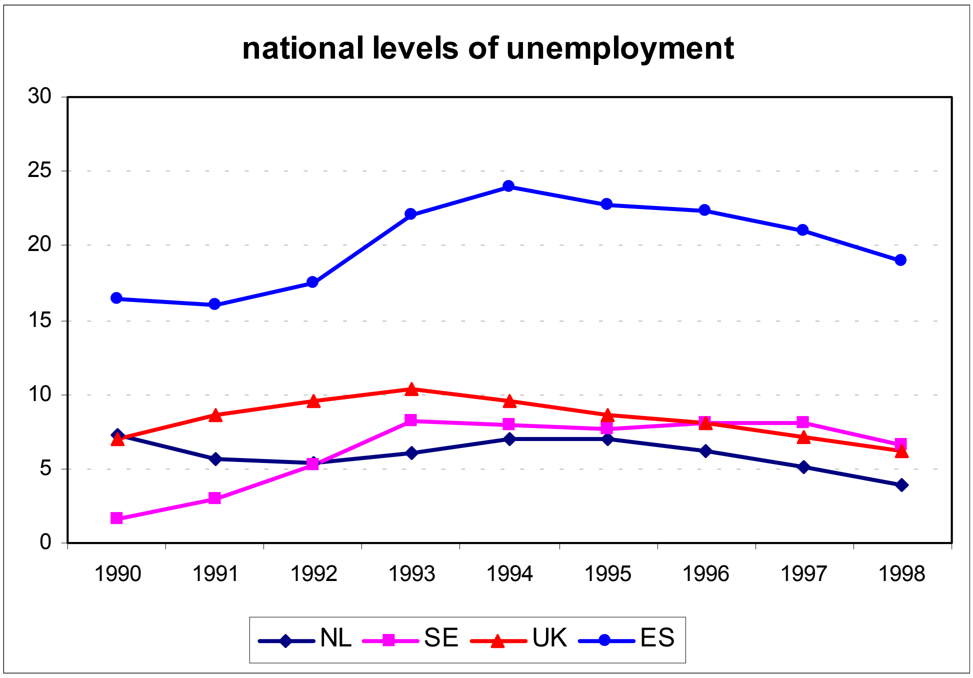


Table 6.4: Correlations between national values and regional disparities(CVs) for GDP per capita and unemployment, 1990-1998

	Netherlands	Sweden	United Kingdom	Spain
GDP per capita	0.89	0.78	0.94	0.87
Unemployment	-0.62	-0.86	-0.86	-0.99

Looking at unemployment we may conclude that in times of rising unemployment, regional differences converge with the largest increases in unemployment in those regions with the lowest levels in the beginning. In times of declining unemployment divergence will take place, but not necessarily in favour of the lowest level regions. The above results are remarkable, as usually regional disparities are known to widen during times of economic recession.

Notwithstanding these developments in disparities, we have to mention that the overall ranking of the regions remains more or less stable over the entire period. For Sweden, the United Kingdom and Spain, for GDP per capita all correlations between years were above 0.9 and for unemployment above 0.8. Relationships for the Netherlands were somewhat lower, but still substantial: 0.85 for GDP per capita and 0.68 for unemployment.

To sum up, we may derive the following implications for internationally consistent scenarios:

- for the case study countries higher growth levels in GDP per capita and decreasing unemployment will coincide with increasing regional disparities, while lower growth levels in GDP per capita and increasing unemployment will coincide with decreasing regional disparities;
- region-specific trend analyses are needed to decide which regions will profit (or suffer) most of economic improvements (or deterioration); and
- for the remaining EU countries, country-specific analyses have to be carried out to investigate whether these results could be generalised

6.5 Data available in New Cronos

To what extent the models tested in this study could be applied in the other EU countries is largely dependent on the data available. In the current section, therefore, an overview will be given of the NUTS 2 internal migration data available in Eurostat's database New Cronos, Where information is missing in New Cronos, it will be indicated whether this could be supplemented by using data of the NSOs. Some comments will be made regarding possible methods to be used in case of missing data. Finally, some remarks will be given on the availability in New Cronos of the explanatory variables GDP per capita, unemployment and population density. It should be stressed here that only data availability has been taken into account. The quality of the data has not been evaluated.

In Table 6.5, characteristics of internal migration data available in New Cronos are summarised:

Table 6.5: Internal migration NUTS 2 data available in New Cronos, EU15

Country (number of NUTS 2 regions)	Internal migration	
	Origin / Destination	O x D
Austria (9)	SA (0-4, 5-9, …, 80-84, 85+) 1992-1999 (1995 missing)	1996-1998 S: 1999 NSO: SA 1996-2002
Belgium (11)	SA (0-4, 5-9, …, 84-89, 90+) 1990-1999 (1997 missing)	1975-1989 (8 regions) S: 1990-1999 (1996-1997 missing)
Finland (6)	SA (0-4, 5-9, …, 85-89, 90+) 1990-1999	1979-1986 (4 regions) S 1987-1999 NSO: SA 1987-2002
France (22+4[a])	Not available	Not available
Germany (40)	A (0-17, 18-24, 25-29,30-49, 50-64, 65+) 1991-1994	1975-1990 (11 regions) S 1987-1999 NSO: SA 1991-1999
Greece (13)	Not available	Not available
Italy (20)	SA (0-4, 5-9, …, 80-84, 85+) 1990-1997	1975-1989 (1981 missing) S 1990-1996 NSO: SA 1952-1999
Netherlands (12)	SA (0-4, 5-9, …, 80-84, 85+) 1990-1999	1975-1985 (9 regions) 1986-1989 S 1990-1999 NSO: SA 1970-2002
Portugal (7)	Not available	1985-1992 (rounded numbers) S 1990 (inconsistent with totals)
Spain (18)	SA (0-4, 5-9, …, 80-84, 85+) 1990-1999 (1995 missing)	1979-1989 S 1990-1999
Sweden (8)	SA (0-4, 5-9, …, 80-84, 85+) 1990-1999	1980-1989 (6 regions) S 1990-1996 (6 regions) S 1997-1999 NSO: SA 1968-2002
United Kingdom (37)	SA (0-4, 5-9, …, 85-89, 90+) 1990-1998 (1995 incomplete)	1979-1996 (rounded totals for 6 Nuts 1 regions)

SA classification by sex (S) and age (A)
S classification by sex, but not by age
A classification by age, but not by sex
[a] 22 mainland regions + 4 Départements d'Outre-Mer

To follow the REGMIG methodology, ideally data are needed on age and sex specific origin-destination flows. As can be seen in Table 6.5, only for six out of 12 countries (Austria, Finland, Germany for a limited number of age groups, Italy, the Netherlands, and Sweden) all data are available either through Eurostat or the NSOs. For the other countries, it remains to be seen to what extent data will be available and accessible. For Belgium, no age-specific origin-destination information is available. For France, information on migration comes from the population census. At INED highly advanced techniques of processing and analysing census data are available, and for the last round of Eurostat scenarios, data for France were provided by INED. For Portugal and Greece no age-specific data were available at all. Internal migration data for Spain are available from the Estadisticas Variciones Residentiales, but it is unclear for what period data are available and to what extent they are accessible. For the UK finally, migration data for sub-national areas come from the National Health Service Register at the level of Family Health Service Authorities. These flows have to be aggregated to NUTS 2 regions.

So, although data are available, some gaps will remain and some adjustments will have to be made in case of missing data. In this respect, possible solutions are:

- if OD-information is available, but age-profiles are not: adopt a national age-specific out-migration profile of a 'similar' country and calculate age-specific in-migration using IPF; in this way, the same age profiles are assumed for each region;
- if OD-information and age-specific in- and out-migration data are available, but it is not possible to distinguish between males and females: use the same patterns for both sexes;
- in case the aggregate OD-information is not consistent with age-specific in- and out-migration data: rescale the OD-flows using iterative proportional fitting (IPF) to get consistent information; and
- in case the OD-matrix is not age-specific: either adopt the same OD-pattern for each age group, or assume OD-age interaction using cluster information of a similar country (for instance regions with high density and low unemployment will have relatively higher in-migration of persons aged 20-60 years old).

As shown in chapter 4, socio-economic indicators are widely available in New Cronos, although again some estimations had to be made. Changes in the NUTS classification, however, are a point of major concern. From the start of the use of the NUTS classification in 1988, several administrative changes have taken place, of which the last one came into effect in 1998. The present NUTS classification divides the 15 countries of the European Union into 210 NUTS 2 regions. In the 1998 revision, a completely new division has been established in the United Kingdom, reflecting the reorganisation of local government during the 1995-1998 period. This resulted mainly in a large increase in the number of NUTS 3 regions, but also in some modifications at NUTS 1 and NUTS 2 level. In Germany, there have been several changes at NUTS 3 level, but also some new regions at NUTS 2 level were introduced. In Sweden, in 1998 the number of NUTS 3 regions has been reduced from 24 to 21. Some counties were merged to form larger regions around the second and third cities of the country. At the same time, two municipalities changed from one county to another, which also affected the NUTS 2 classification. As a consequence of changes in the NUTS classification, no complete long-term data series are available for unemployment. Time series on GDP per capita and population density are less hampered, but also for these indicators some missing values have to be supplemented by using data of the NSOs or by estimation.

6.6 Conclusion

Summarising we come to the following steps to be taken to calculate internal migration rates r_{ijast}, which can be used in regional population scenarios:

1. Collect demographic data from NSOs: age and sex specific origin-destination matrix (OD-matrix) for at least five recent years; in case of missing data, estimate missing parts of the matrix.
2. Collect economic data at the regional level: GDP per capita, unemployment rate; use as far as possible internationally consistent socio-economic data from New Cronos; in case of missing data, use data available at NSOs, or, if no data are available, estimate missing values.
3. Collect data on regional population density.
4. Estimate country-specific models for regional out-migration and destination choice.
5. Calculate elasticities based on parameter estimates.
6. Formulate consistent assumptions on the relevant economic indicators.
7. Impose a time trend for overall migration.
8. Implement the country-specific specifications in the REGMIG model.
9. Calculate the internal migration rates $r_{ijast.}$

From this list of steps to be taken, it should be clear that a country-specific implementation of REGMIG is only possible using both New Cronos data as well as additional country-specific information.

7 Summary and conclusions

7.1 Introduction

For European-wide policies internationally consistent population projections for the European Union are preferred over existing nationally compiled projections. For this reason, the European Commission regularly orders a new revision of national and regional population scenarios for the countries of the EU. In between each new set of scenarios, several background studies are commissioned with the aim to improve the methodology to be used in the next set of scenarios. The present report describes the activities within one of these background studies, entitled ‘Study on past and future interregional migration trends and patterns within EU countries – in search for a generally applicable explanatory model’. The objective of this study was to improve the methodology for compiling internationally consistent scenarios of internal migration. The study has been carried out as a joint effort of the Netherlands Interdisciplinary Demographic Institute (NIDI) and the School of Geography of the University of Leeds (SoG). The point of departure was the methodology used in the latest Eurostat scenarios, the so-called EUROPOP1995 model. The modelling approach proposed in the present study has been indicated as the REGMIG model (REGional MIGration).

A major point of criticism of the EUROPOP1995 model was that country- and region-specific information was not taken into account sufficiently and that there was only a weak theoretical underpinning of the scenarios. Scenarios were formulated in terms of convergence and divergence in the demographic parameters, without direct links to socio-economic developments. In REGMIG it has been examined to what extent explanatory variables could be included in the model and whether it was possible (and necessary) to adopt a life course perspective, i.e. to use different age groups in the models. The overall question to be answered was whether we could move from a “one size fits all method” to a “best model for each country”. In answering this question, a number of other issues are relevant as well. They involve the limitations and possibilities of the available data for each country, the inclusion of country-specific issues in the model, and the practical steps to be taken in preparing the projections. Each of these issues will be dealt with in the next sections. In the final sections the main conclusions will be drawn (7.6) and a number of recommendations for further steps in the direction of actual scenario preparations are given (7.7).

7.2 Current practice

A first step in the project was to review the latest information and experiences with models of sub-national population projections and internal migration used in the Member States of the European

Union themselves. Looking at current practice in the EU-countries, we may state that the use of non-demographic variables, especially for the destination choice of migrants is not widespread. Only the Dutch model includes several non-demographic explanatory variables, for instance housing market and labour market variables. This model, however, is a very sophisticated and country-specific system and using this model in a European-wide context does not seem to be a feasible option. In most other countries very simple techniques of setting internal migration hypotheses have been used. With the exception of a very few examples, these techniques may be defined as no migration projection at all, a status quo projection, and a linear trend extrapolation.

In comparing the current practice with the proposed model in this project, we choose a middle position. The approach is inspired by models used in more advanced settings, although the model's structure is less complex. At the same time, the model is more sophisticated than currently used in many countries. The inclusion of non-demographic variables is a novel element, that is only practised in a minority of the countries.

7.3 Modelling results

In the modelling part of the study we followed up on a number of recommendations for improvement of the EUROPOP1995 model. In particular we tried to incorporate two new elements into the models of internal migration:

1. regional (socio-economic) explanatory variables; and
2. a life course perspective, i.e. explicitly include age in the models (do different models apply to different age groups?).

Internal migration was modelled as a two-stage process with out-migration first followed by destination choice, conditional on out-migration. We analysed data of Sweden, the Netherlands and the United Kingdom and used data for Spain to validate the models. At first, internal migration models were estimated for the period 1991-1995. Subsequently, the models were validated by predicting the flows for 1996-1998 using the estimated model coefficients. Demographic models were estimated and predicted as point of reference for the economic models.

The results of the demographic and economic models are broadly similar for Sweden, the Netherlands and the United Kingdom. For out-migration we found broadly similar specifications of the economic models. Unemployment and Gross Regional Product per capita in purchasing power parities (GDP per capita) are significant explanatory variables for out-migration. Slightly different model specifications were needed for the Netherlands. The explanatory value of the models was only limited. There are large differences in migration behaviour between age groups. This is reflected in the model results.

For destination choice GDP per capita was found to be important to model the attractiveness of regions for all three countries. The added explanatory value of GDP per capita, however, is limited. Internal migration processes are relatively stable in terms of age- and gender patterns, and in terms of regional differences and interaction flows. Despite this limited explanatory power of economic variables, in Sweden and the Netherlands models using economic variables and a distance function were better predictors of destination choice probabilities than purely demographic models.

Notwithstanding these broad similarities the coefficients of the variables show a large variation over countries. For the Netherlands for out-migration we needed a different definition of the variables GDP and unemployment (regional differences compared to the national values). When testing the model for Spain GDP per capita turned out to be important as well.

Change over time is only to a limited extent explained by the out-migration models. Separate models for overall intensities show different economic indicators for different EU countries. Consistent (short term) predictions of internal migration intensities, therefore, are best performed by imposing a time trend.

Summarising, we may conclude that interregional migration is a very complex phenomenon heavily reliant on regional characteristics of individual countries. Formulating a general applicable model is not an easy task to perform. NUTS 2 regions across EU countries do differ significantly in terms of area and population size, and many national and regional peculiarities will have implications for the variables that should be included in a general model. Consequently, a "one size fits all model" is not feasible for all European Union countries. Nevertheless, a "one size fits all *approach*", on the other hand *does* seem to be a possible option. The common approach should have different implementations, which are suited to the migration structure and developments, as well as data availability of each individual country. In this approach, scenarios could be formulated using a relatively limited number of common economic indicators for all countries, whilst country-specific implementations (specifications and coefficients) are needed to do justice to the idiosyncrasies of the individual regions and countries. Moreover, even if the explanatory variables may not be very significant and their predictive value might be limited, they could be useful policy-related levers when applying the models for projection. Further research including additional EU-countries, however, is needed to support this assumption.

7.4 EU-wide scenarios

In general, to set up scenarios we can either use a purely demographic model, a purely economic model or a model which is driven by both demographic and economic developments, a so-called mixed model. Given the results of the current study, we propose to use a mixed model, with socio-economic indicators for the level of migration, region specific out-migration and regional attractiveness. For the spatial component of destination choice models we suggest to use historical origin-destination (OD) patterns instead of distance functions.

Explanatory models are not necessarily better for predicting than historical flow patterns. An extra difficulty in this respect is the projection of the explanatory variables themselves. A major asset of explanatory models is the transparancy of the scenarios. Assumptions on future developments of the socio-economic indicators can be formulated using linear trend analyses or by linking demographic scenarios to (existing) economic scenarios. A drawback of European-wide economic scenarios is that most often these scenarios will not provide detailed input for each of the countries and/or regions separately. Therefore, additional assumptions on country- and regional level will be necessary. Alternatively, country-specific economic scenarios could be used, however, this is only part of the solution as it is highly unlikely that these scenarios will be based on a set of consistent assumptions.

Looking at regional differences in economic developments we may conclude that there are considerable differences between the countries in both GDP per capita as well as unemployment. Especially Spain is lagging behind. The largest regional differences are found in the United Kingdom and Spain, which is not surprising given the higher number of NUTS 2 regions of these countries compared with the Netherlands and Sweden. Disparities were especially large for unemployment. While income disparities slightly increased in the 1990s, disparities in unemployment decreased in the beginning of the 1990s but increased again in later years. Combining disparities with national trends we found a positive relationship for GDP per capita and a negative one for unemployment. Thus, increasing levels of GDP per capita and decreasing unemployment coincide with relatively higher levels of regional disparities. This is remarkable, as generally regional disparities are known to widen during times of economic recession. Notwithstanding these developments the overall ranking of the regions remained more or less stable over the entire period.

A practical issue in setting up scenarios is whether we should formulate assumptions for all regions separately, or whether it is more efficient to develop scenarios for a limited number of groups of regions only. We tried several ways of classifying the NUTS 2 regions on European scale or within the individual countries using GDP per capita and unemployment. Unfortunately, differences between countries and regions are too large to find a classification of regions which can be used for a country-specific implementation of the general model. More detailed country-specific information and expert views are needed to come to a meaningful classification of NUTS 2 regions.

To what extent the REGMIG models could be applied at European scale largely depends on the data available. Ideally data are needed on age and sex specific origin-destination flows. These data could be supplied by the National Statistical Offices for most countries, but not for all. For Portugal and Greece for instance no age-specific internal migration data are available at all. Therefore, some gaps will remain and some adjustments will have to be made in case of missing data. This preliminary phase of estimating missing demographic information is a time consuming but necessary phase of scenario-making. Given the detailed information that is available in other countries this task is feasible. Internationally consistent socio-economic indicators are available in New Cronos, however, due to boundary changes in the NUTS classification, again some adjustments will have to be made. Again, this estimation of missing information step is feasible in preparing the scenarios.

7.5 Country-specific developments

An important outcome of the present study is that interregional migration is highly dependent on regional characteristics of individual countries. At several points the significance of country-specific developments has been stressed. During the workshop with representatives of the four casestudy countries Sweden, the Netherlands, the United Kingdom, and Spain, and the additional countries Germany and Italy, several comments have been made on what developments might be important for internal migration scenarios in these specific countries. A summary of these country-specific features will be given below. Although we have to mention that this overview is far from complete, it does illustrate the different points of interest in different countries, which supports a tailor-made approach for each country within a general framework.

Sweden:

- Additional information on sub-national projections for Sweden:
  The Swedish sub-national projections used constant age-specific intensities of internal migration. These constant intensities were based on a set of recent years and not on any one single year. The NUTS 2 level projections are aggregated from projections at NUTS 5 level. A new population simulation system (Regional Applied Projection System) has been developed for use by county level governments. The system enables the counties to change the projection assumptions and run their own scenarios. County level planners took some time to get used to the system but it is now being used as a planning tool. These local projections using the national sub-national model do not have the status of official projections but rather are simply "what-if" scenarios. If the local scenarios were aggregated for the whole of Sweden, it is likely that the projected population by mid-century would be double the official forecast.

Further features:

- Labour mobility as lubricant in economic development: regional population share goals have been dropped and people are not kept in regions whose economic base has declined
- Social changes in big cities through migration: gentrification at the centre; pauperisation in the middle and gentrification at the periphery
- The number of job vacancies is a more direct economic indicator for explaining time trends in overall migration intensities than inflation rate
- Time trends and fluctuations that must be taken into account: home study programmes that depressed young person migration in the 1980s; increased attraction of large cities for young people in the 1990s and a change after 1990 to a softening in the relationship between vacancies and internal migration
- The education level of the labour force was an important variable explaining migration in the Swedish system

Netherlands:

- Housing-construction as important driving factor of interregional migration (quantitatively as well as qualitatively); other regional characteristics: labour market, education, other services
- Population and households as central entity
- Students and young adults move into the urban areas, while young families and older people move to less densely populated areas

United Kingdom:

Important questions related to policy driven scenarios:

- How could the out-migration losses of big cities be stemmed?

- How could the depressed Thames Gateway towns and districts become in-migration attractors to provide for the considerable demand for jobs in the South-East region?
- What measures might reduce the heavy in-migration to boom areas in the South East with their pressure on the limited housing stock and upward impetus to house prices?

Spain:

- Additional information on sub-national projections for Spain:
  In Spain two sets of official sub-national projections are carried out independently: by the national statistical office, INE and by the Regional Statistical Offices. They are not consistent with each other and use different methods. In general, the focus of INE work is on external migration, particularly immigration rather than on internal migration. A new set of sub-national projections are in preparation by INE which will be informed by the results of the 2001 Census.

Further features:

- Some important return migrations to poor regions
- Young people move to the Mediterranean littorals where tourism is the economic driver

Germany:

- Additional information on sub-national projections for Germany:
  The German Official Statistics Office at Wiesbaden carries out a regular series of sub-national projections at NUTS 1 level (the Länder), the 10^{th} of which has recently been published. The Bundesamt für Bauwesen und Raumordnung (BBR), Bonn carries out sub-regional projections which can be aggregated to NUTS 2 level. The BBR carries out regional and sub-regional projections at the Kreise (440 units) (NUTS 4) and NUTS 3 levels. One innovative feature of the projections is the use of cluster analysis to simplify the specification of time trends in key indicators that drive the projection. The cluster analysis was carried out separately for fertility rates, mortality rates and internal out-migration rates. Trends and scenarios in the development of fertility, mortality and out-migration rates were developed for each cluster rather than for all NUTS 3 or NUTS 4 regions. A second innovative feature of the projections was the adjustment of the OD internal migration matrix to reflect external knowledge about the immigration and subsequent internal migration of *Aussiedler* (immigrants of German descent from central and eastern Europe and former Soviet republics such as Kazakhstan).

Further features:

- Decreasing mobility and migration distances: perhaps indicating a lack of flexibility in responding to regional economic change
- Continuing regional shifts from East to West and from North to South, though in a step by step fashion rather than over long distances
- Continuing suburbanization, which is a family-led process leading to a family type gradient from non-family, singles dominated centres of cities to family dominated suburbs.
- German sub-national projections take into account plans for the operation of asylum seeker centres. The cells in the O-D matrix are adjusted directly to allow for such plans and immigrant relocation after arrival

Italy:
- Migration exchanges are stable and net internal migration shifts are marginal
- Student migration is not important
- Young people migrate on marriage
- Little retirement migration

7.6 Main conclusions

Below six main conclusions are drawn, based on the central research questions of this project.

1. *The main objective of this study was to improve the methodology for making internal migration projections in EU countries.* This was done in two ways. First, since internal migration is driven by many non-demographic (economic, geographical, other) variables, this project investigated whether using non-demographic variables in addition to demographic variables are able to improve internal migration forecasts. Second, a life course perspective was adopted, in which the migration model structure is different for age groups.

2. *Is there added value of non-demographic variables?* The results indicate that non-demographic variables give limited added value in explaining migration trends, although in some countries (NL, SE) they give better short-term predictions. However, this conclusion leaves aside the issue of the predictability of non-demographic variables, such as GDP or unemployment. For scenario-making this should not be a large problem though. Thus, using non-demographic information is worthwhile for constructing internal migration scenarios. This also provides policy-makers with explicit levers to change migration assumptions.

3. *Is there added value of the life course approach?* The results indicate that the explanatory value of non-demographic variables is very different across age categories, as well as (although to a much lesser extent) for men and women, with largest differences between the young and mobile age groups and the elderly. Leaving out the age/sex dimension significantly reduces the explanatory value of the models. Therefore, aggregation of migration over all age categories is not warranted. The life course is an important dimension and should be taken into account.

4. *One size fits all?* The model structures are not only different for each age/sex category, but also vary across countries. Therefore, models should be specified separately for each country. Nevertheless, a *one-approach fits all* is necessary, with a common model form at the basis of different implementations for each country. This country-specific implementation should take into account the specificities and data-availability of each country.

5. *Are data a problem for this approach?* In answering this questions we have to make a distinction between demographic and non-demographic data. With respect to demographic information this approach can be implemented based on available data that is to some degree present in New Cronos, and additional data from the National Statistical Offices. For some countries an additional phase of estimating missing information is necessary. Taken together, the demographic data are not a constraint for implementing this method. With respect to non-

demographic information some additional data estimation is also necessary, for a number of countries, but here again this should not obstruct the implementation of this method.

6. *How should country-specific issues be taken into account?* It is clear from the project results that migration processes are very different across countries. There are a number of ways to deal with the issue of country-specific elements. First, the regional classification is largely different across countries. One way to reduce this heterogeneity is to make a classification of regions, based on regional characteristics. Results indicate that a purely statistical approach is not sufficient, and that country-specific information and expert opinions are necessary. For each regional cluster separate assumptions should be made. Second, additional assumptions may be necessary for individual countries that reflect specific developments in internal migration. One example is the system of regional dispersion of asylum seekers in individual countries.

7.7 Recommendations for future research

In this final section of the report we present some general remarks and recommendations for future research.

- *Analysis of residuals*
  Although in general both the demographic as well as the economic models tested in this study resulted in satisfactory predictions of migration patterns, there still was some systematic under- or overestimation. Examination of the spatial patterns of errors between observed and predicted patterns might give some more clarification. By mapping the largest errors, for instance, some further explanatory variables might be suggested.

- *Feedback mechanism of economy and demography*
  A problem in using economic indicators in demographic scenarios is that some of them, for instance unemployment, depend on the size and the composition of the population, which makes the argument circular. In order to solve this problem a demo-economic model has to be used in which both the demographic as well as the economic variables are endogenous to the model. Within the setting of REGMIG, however, this problem cannot be solved.

- *Backwards harmonization of regional classifications*
  One of the problems with long term time series on internal migration is the lack of reconstruction of past time series when regional boundary changes occurred. A system of georeferencing might be a solution. Statistics Finland, for instance can produce aggregations of register based information such as internal migration using suitable look-up tables between georeferences and regional boundaries. In this way they can reconstruct for the past time series based on current regional definitions. In the United Kingdom the National Statistical Offices (ONS, GROS, NISRA) have developed a georeferencing system based on the unit postcode, which is the finest location reference in the system of postal addresses and contain 10-15 residences typically. Unfortunately, the statistical infrastructures in many EU Member States may not be good enough to support this backwards harmonisation of demographic data. A

recommendation to create the conditions to set up such a system might be a first step in the right direction.

- *Housing market variables and commuting*
  The explanatory variables used in the migration models are mainly labour market related. However, housing market variables together with trade-offs between migration and commuting are also important. Housing construction may play a part in explaining the pace of population change in growing regions, while also house prices may have important influences on migration. Although this argument was fully appreciated, no variables relating to the housing market are consistently available across all EU Member States.

- *Links between asylum migration and regional distribution of internal migrants*
  In Europe, in the first half of the 1990s a large part of international migration consisted of asylum migration. This might have had consequences for the internal migration patterns of recent immigrants, especially if policy measures had dispersed immigrants to a wide variety of locations. For instance in Sweden immigrants had subsequently concentrated in larger cities. In Germany, links between asylum migration and internal migration patterns were explicitly taken into account in projection models. Impacts at NUTS 2 level, however, may be limited, either because immigrants make up only a small part of all internal migrants, or because internal migrations following international migration may often take place within the same NUTS 2 regions.

- *Net effect of internal migration*
  For the overall result of the projections, the out-migration and destination patterns are not of direct relevance; what is important is the combination of out- and in-migration, i.e. net internal migration. As it is net-migration that will shift the population in the projections, net-migration could be used as a measure of model success as well. However, for the transparency of the projections, the underlying processes of out-migration and destination choice are at least as important.

- *Regional dynamics*
  Regions are not simple attractive regions with high in-migration and low out-migration levels, or unattractive regions with low in-migration and high out-migration. Often, high out-migration regions are also high in-migration regions. Further research is needed to answer the question of whether it is possible to improve predictions by taking into account the high correlation between out- and in-migration.

- *The intergenerational conflict hypotheses*
  Looking at origin-destination patterns by age gives the opportunity to test a set of intergenerational conflict hypotheses. For instance, is out-migration of the older population (60+) triggered by the in-migration of young persons (15-19, 20-29) with different and incompatible lifestyles? Although these hypotheses have not been taken into account in the current study, age turned out to be an important selective influence on migration. Age-specific infrastructure variables (e.g. university places, health care facilities, or environmental variables)

may also be important here. Further research in this field is recommended for a better understanding of all internal migration dynamics.

References

Amrhein, C.G. and Flowerdew, R. (1992) The effect of data aggregation on a Poisson regression model of Canadian migration, *Environment and Planning A*, 24: 1381-1391.

Armitage, R. (1986) Population projections for English local authority areas, *Population Trends*, 43(Spring): 31-40.

Baydar, N. (1983) Analysis of the temporal stability of migration in the context of multi-regional forecasting, *Working Paper No 38,* Netherlands Interuniversity Demographic Institute, Voorburg.

Barff, R. (1990) The migration response to the economic turnaround in New England, *Environment and Planning, 22*: 1497-1516.

Bates, J.J. and Bracken, I. (1987) Estimation of migration profiles for local authority areas in England, 1971-1981, *Environment and Planning A*, 19(4): 521-35.

Bengtsson, T. and Johansson, M. (1995) The new migration transition – the case of post-industrial Sweden, *Demography, Economy and Welfare*, 10: 217-235.

Birg, H. (1983) On the interactions of job creation, migration, and natural population increase within the framework of a dynamic demo-economic model, *Environment and Planning A, 14*: 1141-1153.

Bohara, A.K. and Krieg, R.G. (1996) A zero-inflated Poisson model of migration frequency*, International Regional Science Review*, 19(3): 211-22.

Bowley, A.L. (1924) Births and population change in Great Britain, *The Economic Journal,* 34: 188-192.

Boden, P., Stillwell, J.C.H. and Rees, P.H. (1991) Internal migration projection in England: the OPCS/DOE model examined, Chapter 14 in Stillwell, J.C.H. and Congdon, P. (eds.) *Migration Models: Micro and Macro Perspectives,* Belhaven Press, London, pp. 262-286.

Boyle, P. (1993) Modelling the relationship between tenure and migration in England and Wales, *Transactions, Institute of British Geographers*, 18: 359-376.

Boyle, P. J., Flowerdew, R. and Shen, J. (1998) Analysing local-area migration using data from the 1991 Census: the importance of housing growth and tenure, *Regional Studies,* 32(2): 113-132.

Bracken, I. and Bates, J.J. (1983) Analysis of gross migration profiles in England and Wales: some developments in classification, *Environment and Planning A*, 15: 343-55.

Cadwallader, M. (1989) A conceptual framework for analysing migration behaviour in the developed world, *Progress in Human Geography*, 13(4): 494-511.

Champion, A.G., Bramley, G., Fotheringham, A.S., Macgill, J. and Rees, P.H. (2002) A migration modelling system to support government decision-making, in Stillwell, J. and Geertman, S. (eds.) *Planning Support Systems in Practice*, Springer Verlag, Berlin, pp.257-278.

Champion, A.G., Fotheringham, A.S., Rees, P., Boyle, P.H. and Stillwell, J.C.H. (1998) The *Determinants of Migration Flows in England: a Review of Existing Data and Evidence*, Department of Geography, University of Newcastle, for the Department of Environment, Transport and the Regions. Available as http://www.geog.leeds.ac.uk/publications/DeterminantsOfMigration/report.pdf

Congdon, P. (1991) An application of general linear modelling to migration in London and the South East, Chapter 7 in Stillwell, J.C.H. and Congdon, P. (eds.) *Migration Models: Macro and Micro Perspectives*, Belhaven Press, London, pp. 113-136.

Courgeau, D. (1973) Migrations et découpage du territoire, *Population*, 28, 3: 511-537.

Courgeau, D. (1995) Migration theories and behavioural models, *International Journal of Population Geography*, 1: 19-27.

Davies, R. and Guy, C. (1987) The statistical modelling of flow data when the Poisson assumption is violated, *Geographical Analysis*, 19: 300-314.

De Gans, H.A. (1999) *Population Forecasting 1895-1945: The Transition to Modernity,* Kluwer Academic, Dordrecht.

De Mooij, R. and Tang, P. (2003) *Four futures of Europe*, CPB document, The Hague.

Devis, T. (1983) People changing address: 1971 and 1981, *Population Trends*, 32: 15-20.

Dunford, M., and Fielding, A.J. (1997) Greater London, the south-east region and the wider Britain: metropolitan polarization, uneven development and inter-regional migration. In: H.H. Blotevogel, and A.J. Fielding (eds.) *People, Jobs and Mobility in the New Europe.* John Wiley & Sons, Chichester, pp. 247-276.

Easterlin, R.A. (1980) *Birth and Fortune. The Impact of Numbers on Personal Welfare*, Basic Books, New York.

European Communities (1999) Regions, nomenclature of territorial units for statistics – NUTS.

Feeney, G. (1973) Two models for multiregional population dynamics, *Environment and Planning A*, 5: 31-43.

Fielding, A.J. (1990) Migration and social mobility: South East England as an escalator region, *Regional Studies*, 26: 1-15.

Fielding, A.J. (1992) Migration and social change, in Stillwell, J.C.H., Rees, P.H. and Boden, P. (eds.) *Migration Processes and Patterns Volume 2: Population Redistribution in the United Kingdom*, Belhaven, London.

Flowerdew, R. (1991) Poisson regression modelling of migration, in Stillwell, J.C.H. and Congdon, P. (eds.) *Migration Models: Macro and Micro Approaches*, Belhaven Press, London, pp. 92-112.

Flowerdew, R. and Lovett, A. (1988) Fitting constrained Poisson regression models to interurban migration flows. *Geographical Analysis,* 20: 297-307.

Fotheringham, A.S. (1983) A new set of spatial interaction models: the theory of competing destinations, *Environment and Planning A*, 15: 15-36.

Fotheringham, A.S. (1986) Modelling hierarchical destination choice, *Environment and Planning A*, 18: 401-418.

Fotheringham, A.S. (1991) Migration and spatial structure: the development of the competing destinations model, in Stillwell, J. and Congdon, P. (eds.) *Migration Models: Macro and Micro Approaches,* Belhaven, London, pp. 57-72.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Wiley, Chichester.

Fotheringham, A.S., Nakaya, T, Yano, K., Openshaw, S. and Ishikawa, Y. (2001) Hierarchical destination choice and spatial interaction modelling: a simulation experiment, *Environment and Planning A*, 33: 901-920.

Fotheringham, A.S. and O'Kelly, M.E. (1989) *Spatial Interaction Models: Formulations and Applications*, Kluwer Academic, Dordrecht.

Fotheringham, A.S., Rees, P., Champion, A., Kalogiou, S. and Tremayne A.R. (2004) The development of a migration model for England and Wales I: overview and modelling out-migration, *Environment and Planning A*, 36 (forthcoming).

Hughes, G. and McCormick, B. (1981) Do council housing policies reduce migration between regions?, *Economic Journal*, 91: 919-937.

Hughes, G. and McCormick, B. (1985) Migration intentions in the UK. Which households want to migrate and which succeed?, *Economic Journal, Conference Supplement,* 95: 113-123.

Hughes, G. and McCormick, B. (1987) Housing markets, unemployment and labour market flexibility in the UK, *European Economic Review*, 31: 615-645.

Hughes, G. and McCormick, B. (1989) Does migration reduce differentials in regional unemployment rates?, in van Dijk, J., Folmer, H., Herzog, Jr, H.W. and Schlottmann, A.M. (eds.) *Migration and Labour Market Adjustment*, Kluwer, Dordrecht, pp. 85-108.

INE (1993) Indicadores Sociales 1991.

Kemper, F.J. (1997), Internal migration and the business cycle: the example of West Germany. In: H.H. Blotevogel, and A.J. Fielding (eds.) *People, Jobs and Mobility in the New Europe.* John Wiley & Sons: Chichester, pp. 227-245.

Kitsul, P. and Philipov, D. (1981) The one-year/five-year migration problem, in Rogers, A. (ed.), *Advances in multiregional demography,* RR-81-6, IIASA, Laxenburg.

Kupiszewski, M. and Kupiszewska, D. (2003) Internal migration component in subnational population projections in member states of the European Union, *Working paper Central European Forum of Migration Research*, Warsaw, Poland.

Lee, E S.(1966) Theory of migration, in Hauser, P.H. and Duncan, O.D. (eds.) *The Study of Population,* University of Chicago Press, Chicago, pp. 282-297.

Lee, E. (1969) A theory of migration, in Jackson, J.A. (ed.) *Migration*, Cambridge University Press.

Lejour, A. (2003) *Quantifying Four Scenarios for Europe*, CPB document 38, The Hague.

Leslie, P.H. (1945) On the use of matrices in certain population mathematics, *Biometrika*, 33: 183-212.

Leslie, P.H. (1948) Some further notes on the use of matrices in population mathematics, *Biometrika,* 35: 145-213.

Liaw, K.L. and Ledent, J. (1987) Nested logit model and maximum quasi-likelihood method: a flexible methodology for analysing interregional migration patterns, *Regional Science and Urban Economics*, 17: 67-88.

Lowry, I. (1966) *Migration and Metropolitan Growth: Two Analytical Reports*, Chandler, San Francisco.

Maier, G. and Weiss, P. (1991) The discrete choice approach to migration modelling, Chapter 2 in Stillwell, J.C.H. and Congdon, P. (eds.) *Migration Models: Macro and Macro Approaches*, Belhaven Press, London, pp.17-33.

Milne, W.J. (1993) Macro-economic influences on migration. *Regional Studies, 27 (4)*: 365-373.

Mønnesland, J. (1997) Population redistribution in the sparsely populated periphery: the Nordic case. in Blotevogel, H.H. and Fielding, A.J. (eds.), *People, Jobs and Mobility in the New Europe.* John Wiley & Sons: Chichester, pp. 153-167.

Myrdal, G. (1957) *Rich Lands and Poor*, Harper and Row, New York.

Nakaya, T. (2001) Local spatial interaction modelling based on the geographically weighted regression approach, *GeoJournal*, 53: 347-358.

Öberg, S. (1997) Theories on inter-regional migration: an overview. in Blotevogel, H.H. and Fielding, A.J. (eds.), *People, Jobs and Mobility in the New Europe.* John Wiley & Sons: Chichester, pp. 23-48.

ODPM (2002) *Development of a Migration Model*, Report prepared by the University of Newcastle upon Tyne, the University of Leeds, and the Greater London Authority/London Research Centre, ODPM, London. Available as http://www.housing.odpm.gov.uk/information/migration/pdf/model.pdf.

ONS (2002) Report: internal migration estimates for local and health authorities in England and Wales, 2001, *Population Trends,* 109: 87-99.

Owen, D. and Green, A. (1992) Migration patterns and trends, in Champion, A.G. and Fielding, A.J. (eds.) *Migration Processes and Patterns Volume 1: Research Progress and Prospects*, Belhaven, London.

Pandit, K. (1997) Cohort and period effects in US migration: how demographic and economic cycles influence the migration schedule, *Annals of the Association of American Geographers*, 87: 439-450.

Plane, D.A. (1982) An information theoretic approach to the estimation of migration flows, *Journal of Regional Science*, 22: 441-456.

Plane, D.A. (1992) Demographic influences on migration, *Regional Studies*, 27(4): 375-383.

Plane, D.A. and Rogerson, P.A. (1986) Dynamic flow modelling with inter-regional dependency effects: an application to structural change in the US migration system, *Demography*, 23: 91-104.

Plane, D.A. and Rogerson, P.A. (1991) Tracking the babyboom, the baby bust, and the echo generations: how age composition regulates US migration, *Professional Geographer*, 43: 416-30.

Plane, D.A. and Rogerson, P.A. (1994) *The Geographical Analysis of Population with Applications to Planning and Business,* Wiley, New York.

Ravenstein, E.G. (1885) The laws of migration, *Journal of the Royal Statistical Society*, 48(II): 167-227.

Rees, P. (1977) The measurement of migration from census data and other sources. *Environment and Planning A*, 9: 247-272.

Rees, P.H. (1981) Accounts-based models for multi-regional population analysis: methods, program and user's manual, *Working Paper 295/Computer Manual 13,* School of Geography, University of Leeds, Leeds.

Rees, P.H. (1984) Spatial population analysis using movement data and accounting methods: theory, models, the MOVE program and examples, *Working Paper 404/Computer Manual 23*, School of Geography, University of Leeds, Leeds.

Rees, P.H., Fotheringham, A.S. and Champion, A.G. (2004) Modelling migration for policy analysis, Chapter 14 in Stillwell, J. and Clarke, G. (eds.) *Applied GIS and Spatial Analysis*, Wiley, Chichester, pp. 259-296 (forthcoming).

Rees, P. and Kupiszewski, M. (1999a), *Internal Migration and Regional Population Dynamics in Europe: a Synthesis*. Population Studies No.32, Council of Europe Publishing, Strasbourg, pp. 113.

Rees, P. and Kupiszewski, M. (1999b) Internal migration: What data are available in Europe? *Journal of Official Statistics,* 15(4): 551-586.

Rees, P., Kupiszewski, M., Eyre, H., Wilson, T. and Durham, H. (2001) The evaluation of regional population projections for the European Union, *Eurostat Working Paper, Population and Social Conditions* 3/2001/E/no 9.

Rees, P.H., Stillwell, J.C.H., Boden, P., Duley, C., Ramsden, C., Bristow, M., Patel, I. and Rees, G. (1990) *SWIS: An Information and Projection System for Ward Populations and Housing for the City of Swansea, Manual and Report*, Report for GMAP Ltd, University of Leeds, Leeds.

Rees, P., van Imhoff, E., Durham, H., Kupiszewski, M., and Smith, D. (1998) *Internal Migration and Regional Population Dynamics in the Netherlands.* NIDI report no. 55, NIDI, The Hague.

Rees, P. and Willekens, F. (1986) Data and accounts, Chapter 2 in Rogers, A. and Willekens, F. (eds) *Migration and Settlement: a Multiregional Comparative Study*, Reidel, Dordrecht.

Rees, P.H. and Wilson, A.G. (1973) Accounts and models for spatial demographic analysis 1: aggregate population, *Environment and Planning*, 5: 61-90.

Rees, P.H. and Wilson, A.G. (1975) A comparison of available models of population change, *Regional Studies*, 9: 39-61

Rees, P.H. and Wilson, A.G. (1977) *Spatial Population Analysis*, Edward Arnold, London.

Rogers, A. (1966) Matrix methods of population analysis, *Journal of the American Institute of Planners,* 32: 40-44.

Rogers, A. (1967) Matrix analysis of interregional population growth and distribution, *Papers of the Regional Science Association,* 18: 17-196.

Rogers, A. (1968) *Matrix Analysis of Interregional Population Growth and Distribution*, University of California Press, Berkeley.

Rogers, A. (1973) The multi-regional life table, *Journal of Mathematical Sociology*, 3: 127-137.

Rogers, A. (1990) Requiem for the net migrant, *Geographical Analysis*, 22: 283-300.

Rogers, A. (1995) *Multiregional Demography,* Wiley, London.

Rogers, A. and Castro, L.J. (1981) *Model Migration Schedules, RR-81-30*, International Institute of Applied Systems Analysis, Laxenburg, Austria.

Rogers, A., Raquillet, R. and Castro, L.J. (1978) Model migration schedules and their applications, *Environment and Planning A,* 10: 475-502.

Rogers, A. and Planck, F. (1984) MODEL: a general program for estimation of parameterized model schedules of fertility, mortality, migration and marital status and labour force status transitions, *Working Paper 83-102*, IIASA, Laxenburg, Austria.

Rogers, A. and Willekens, F. (eds.) (1986) *Migration and Settlement: A Multiregional Comparative Study,* Reidel, Dordecht.

Rossi, P. (1955) *Why Families Move: A Study in the Social Psychology of Urban Residential Mobility*, Free Press, Glencoe, Illinois.

Rothenberg, J. (1977) On the microeconomics of internal migration, in Brown, A.A. and Neuberger, E. (eds.) *Internal Migration: A Comparative Perspective*, Academic Press, New York, pp. 183-208.

Scholten, H.J. and van Wissen, L.J.G. (1985) A comparison of the loglinear interaction model with other spatial interaction models. In: P. Nijkamp, H. Leitner and N. Wrigley (eds.) *Measuring the Unmeasurable*, Martinus Nijhoff Publishers, Dordrecht, pp. 177-196.

Sjaastad, L. (1960) The relationship between migration and income in the United States*, Papers of the Regional Science Association*, 6: 37-64.

Stambøl, L. (1991) Migration projection in Norway: a regional demographic-economic model, Chapter 15 in Stillwell, J.C.H. and Congdon, P. (eds.) *Migration Models: Macro and Micro Perspectives*, Belhaven Press, London, pp. 287-308.

Stillwell, J.C.H. (1978) Interzonal migration: some historical tests of spatial interaction models, *Environment and Planning A*, 10: 1187-1200.

Stillwell, J.C.H. (1984) IMP: a program for inter-area migration analysis and projection: user's manual (revised), *Computer Manual 12*, School of Geography, University of Leeds, Leeds.

Stillwell, J.C.H. (1991) Spatial interaction models and the propensity to migrate over distance, Chapter 3 in Stillwell, J.C.H. and Congdon, P. (eds.) *Migration Models: Micro and Macro Approaches*, Belhaven Press, London, pp. 34-56.

Stillwell, J.C.H. (1994) Monitoring intercensal migration in the United Kingdom, *Environment and Planning A*, 26: 1711-1730.

Stillwell, J.C.H., Bell, M., Blake, M., Duke-Williams, O. and Rees, P.H. (2001) A comparison of net-migration flows and migration effectiveness in Australia and Britain: Part 2, Age-related migration patterns, *Journal of Population Research*, 18(1): 19-39.

Stillwell, J.C.H. and Congdon, P. (1991) (eds.) *Migration Models: Macro and Macro Approaches*, Belhaven Press, London.

Stillwell, J.C.H., Duke-Williams, O. and Rees, P.H. (1995) Time series migration in Britain: the context for 1991 Census analysis, *Papers in Regional Science: Journal of the Regional Science Association*, 74(4): 341-359.

Stillwell, J,C.H. Rees, P.H. and Boden, P. (eds.) (1992) *Migration Processes and Patterns. Volume 2, Population Redistribution in the United Kingdom,* Belhaven, London.

Thomas, D.S. (1938) *Research Memorandum of Migration Differentials*, Social Science Research Council, New York.

Van Fulpen, H. (1985) *Volkshuisvesting in Demografisch en Economisch Perspectief.* Sociaal en Cultureel Planbureau, Staatsuitgeverij, 's-Gravenhage.

Van der Gaag, N., van Imhoff, E. and van Wissen, L. (1997a), Regional population projections in the countries of the European Economic Area, update of the 1992-Questionnaire. *NIDI Working paper* no.: 97/1, The Hague

Van der Gaag, N., van Imhoff, E. and van Wissen, L. (1997b) Long-term internal migration scenarios for the countries of the European Union, *Eurostat Working paper no.:* E4/1997-5.

Van der Gaag, N., van Imhoff E. and van Wissen, L. (2000) Internal migration scenarios and regional population projections for the European Union, *International Journal of Population Geography*, vol. 6(1):. 1-19.

Van der Gaag, N. and van Wissen, L. (2001), Economic developments and internal migration propensities. Paper presented at the European Population Conference 2001, Helsinki, 7-9 June.

Vanhove, N. (1999) *Regional Policy: A European Approach* (third edition), Ashgate, Aldershot.

Van Imhoff, E., van der Gaag, N., van Wissen, L. and Rees, P.H. (1997) The selection of internal migration models for European regions, *International Journal of Population Geography*, 3: 137-159.

Van Imhoff, E., van Wissen, L. and Spiess, K. (1994), *Regional Population Projections in the countries of the European Economic Area*, NIDI CBGS Publications no. 31, The Hague, Swets and Zeitlinger.

Van Solinge, H., van Dalen, H., Dykstra, P., van Imhoff, E., Moors, H. and van Wissen, L. (1998) *Population, Labour and Social Protection in the European Union: Dilemmas and Prospects.* NIDI report 52, The Hague

Van Wissen, L., Nijkamp, P. and Rima, A. (1991) The functioning of the housing market in Amsterdam. In: Hårsman, B.and Quigley, J.M. (eds) *Housing Markets and Housing Institutions: An International Comparison*, Kluwer Academic Publishers, Boston, pp. 113-155.

Van Wissen, L.J.G. and Rima, A. (1988) *Modelling Urban Housing Market Dynamics,* Studies in Regional Science and Urban Economics, North Holland-Amsterdam.

Warnes, A. (1992) Migration and the life course, in Champion, A.G. and Fielding, A.J. (eds.) *Migration Processes and Patterns Volume 1: Research Progress and Prospects*, Belhaven, London.

Whelpton, P.K. (1936) An empirical method of calculating future population, *Journal of the American Statistical Association*, 31: 457-473.

White, S.E. (1980) A philosophical dichotomy in migration research, *Professional Geographer*, 32: 6-13.

Willekens, F. and Drewe, P. (1984) A multi-regional model for regional demographic projection, in ter Heide, H. and Willekens, F. J. (eds.) *Demographic Research and Spatial Policy: The Dutch Experience*, Academic Press, London, pp. 309-334.

Wilson, A.G. (1967) A statistical theory of spatial distribution models, *Transportation Research,* 1: 253-269.

Wilson, A.G. (1970) *Entropy in Urban and Regional Modelling*, Pion, London.

Wilson, A.G. (1972) *Papers in Urban and Regional Analysis*, Pion, London.

Wilson, A.G. and Rees, P.H. (1974) Accounts and models for spatial demographic analysis 2: age-sex disaggregated populations, *Environment and Planning A*, 6: 101-116.

Wilson, T. (2001) A new subnational population projection model for the United Kingdom, *Unpublished PhD Thesis*, School of Geography, University of Leeds, Leeds.

Woods, R.I. and Rees, P. (ed.), (1986) *Population Structures and Models*, Allen and Unwin, London.

Yano, K., Nakaya, T., Fotheringham, A.S., Openshaw, S. and Ishikawa, Y. (2003) A comparison of migration behaviour in Japan and Britain using spatial interaction models, International Journal of Population Geography, 9(5): 419-431.

Zipf, G.K. (1946) The P1P2/D hypothesis: on intercity movement of persons, *American Sociological Review*, 11: 677-86.