"An investigation into the hierarchical nature of internal migration destination choice in Great Britain"

David Atkins

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

September 2006

Abstract

This research investigates the psychological basis of the process of migration destination choice. Specifically, it attempts to demonstrate that this process is hierarchical in nature, meaning that migrants group potential destinations into clusters or regions and examine the characteristics of groups of destinations before comparing specific locations within preferred regions.

An examination of the variation in this hierarchical behaviour between migrants from different origins also provides valuable insights into the nature and role of cognitive space: that is, each individual's internal representation of the world around them.

Four different techniques are used to illustrate hierarchical aspects of migrant's decision-making: the competing destinations model; two variants of the nested logit model; and a novel hybrid model. Hierarchical behaviour is demonstrated through comparison of these models with the traditional 'flat-processing' migration model. Migration data used in this research covers moves within Great Britain and is derived from the 1991 UK Census of Population with several other sources providing information describing the various potential migration destinations.

This research will contribute both theoretically and methodologically to the field of migration research. It is novel in its hierarchical approach to modelling the UK internal migration destination choice process, and also in its algorithms for generating the discrete and probabilistic regionalisations of space that are used to represent migrants' hierarchical destination choice sets. It is hoped this work will be of practical benefit to the population and migration forecasting community by providing the basis for an inherently more accurate hierarchical approach to migration destination choice modelling.

Dedication

"Mum and Dad, this one's for you - thanks for everything!"

Table of Contents

List of Tables

List of Figures

List of Maps

List of Equations

Acknowledgements

I would like to thanks the many friends and family who have provided considerable encouragement and tireless support over the extended duration of my study for this doctorate. I am very grateful to my supervisor and friend, Stewart Fotheringham, for his wise guidance and direction, and also his patience and understanding.

I would particularly like to thank my wonderful wife and best friend, Tina Wu. It is no understatement to say that this thesis would never have been completed without her unwavering support. She has motivated me when I didn't think that was possible and has sheltered me from many mundane day-to-day activities to keep me focused on this PhD.

Thank you all so very much!

Chapter One

Is migration destination choice a hierarchical process?

Migration in Great Britain has for many decades been the primary mechanism of population redistribution. As such, the ability to accurately predict migration behaviour is essential to producing reliable predictions of future population patterns. Many of the planning activities of both private and public organizations need to take account of expected future population patterns. However, the techniques currently employed to predict migration have not changed significantly over the last four decades, despite recent advances in the field of spatial interaction modelling.

Two main approaches have been adopted for the analysis of migration behaviour. Behavioural researchers have performed much qualitative analysis primarily using interview and questionnaires to examine the motivations of small numbers of individual migrants. At a more aggregate scale quantitative migration researchers have applied many statistic and mathematic tools to examine patterns in observed migration behaviour, particularly using various types of spatial interaction models to try and predict the population redistribution that results from internal migration. This research attempts to close the gap between these disciplines by using individual-level behavioural principles to inform the derivation and application of a new generation of aggregate-level migration destination choice models.

Traditional migration models have employed a 'flat-processing' approach to the migration destination choice process. Such an approach assumes that each migrant is aware of and has characteristic information describing every potential destination, and that each destination is assessed in relation to all other destinations. Common sense suggests that this is unlikely to be an accurate reflection of the decision-making process underlying the selection of migration destinations, as it assumes vast information storage and processing capability, as well as the unlikely situation that each migrant would have had cause and opportunity to gather information about every possible migration destination.

Real world constraints on individuals' access to and ability to process information describing potential migration destinations suggest that a more selective decision-making process is more likely to underpin migration behaviour. A more efficient selection mechanism would be a spatially hierarchical decision-making process. Using such an approach, a migrant would sequentially compare and select between destination areas at increasingly fine spatial scales, limiting the number of alternatives assessed at each spatial level of the hierarchy to a manageable number. Modern spatial interaction models provide a mechanism for predicting such spatially hierarchical decisions by explicitly including spatial structure into the model specification. However, such models have not yet been widely applied.

The lack of application of such modern techniques to migration analysis is largely due to a belief that these techniques are out of date and based on outmoded assumptions. This neglect is an unfortunate result of modern hierarchical models appearing superficially to be similar in operational terms to the over-simplistic traditional gravity model. However, as will be demonstrated in chapter two, spatial interaction modelling has evolved significantly since the early days of gravity models, such that current criticisms are largely unfounded.

The earliest migration models were proposed by Ravenstein, who, in 1885 published *'The Laws of Migration'* which introduced the principles of 'social physics' as the basis for modelling the movement of migrants (Ravenstein, 1885). Ravenstein's *Laws* were derived purely from examination of observed migrant flows and take no account of individual-level decision-making processes. Ravenstein noted that migration data indicated that more migrants move to closer and larger destination settlements than move to smaller and more distant destinations. This observation prompted Ravenstein to draw parallels between the attractiveness of migration destinations and the gravitational pull of massive objects. This analogy gave rise to the term social physics and caused early spatial interaction models to be coined gravity models.

Today's discrete choice models are based largely upon psychological and geographic theories of individual behaviour, particularly rational utility maximization. However, the operational formulation of these models appears similar in structure to the traditional 'gravity' model introduced by Ravenstein.

This thesis introduces a number of recently developed and novel analytical techniques and demonstrates their application to the modelling of migration destination choice in Great Britain. The derivation of these modern migration models is described and their evolution from traditional models is discussed.

Conclusions are drawn based on the relative goodness of fit of these models. Because the primary goal of this research is to gain insight into the cognitive process of migration destination choice, as opposed to the identification of all the major determinants of migration, the absolute accuracy of the various models is of secondary interest.

Specifically, this research attempts to demonstrate that the destination choice process is inherently hierarchical in nature. To this end, a number of modern techniques of spatial interaction modelling are employed and extended to model the migration destination choice process in a fundamentally hierarchical manner. It is anticipated that comparison of results obtained from the calibration of such hierarchical models and of other traditional migration models will provide new evidence in support of the theory that the migration destination choice process is indeed hierarchical in nature.

General approach

Migration modelling, indeed spatial interaction modelling in general, has often been seen as over simplistic and been criticised for imposing too many unrealistic assumptions upon the migration process. This ambivalence is often due to a lack of widespread appreciation of recent developments in spatial interaction modelling, and is largely a throwback to the earlier generation of empirically derived 'gravity models' of spatial interaction. To help dispel these outdated beliefs, it is beneficial to introduce the development of spatial interaction modelling to date.

Fotheringham described the evolution of spatial interaction modelling as having gone through a number of fairly distinct stages: gravity models based on social physics; entropy-maximizing models based on statistical mechanics; aspatial discrete choice models, from economics / transportation science; and, spatial discrete choice models, incorporating spatial structure of the choice set (Fotheringham, 2000).

The first phase of migration modelling was the gravity models that applied the principles of social physics introduced in Ravenstein's *Laws* (Ravenstein, 1885). Ravenstein's models were derived from examination of empirical migration data rather than any theory of migration behaviour. This analysis was the first to observe the general attractiveness of larger destinations and deterrent effect of greater potential migration distances, which was compared to the gravitational force that any object with mass exerts and the fact that such gravitational attractiveness reduces across distance – hence the naming of his models gravity models.

Wilson was the first to present a theoretical framework for the derivation of spatial interaction models. Wilson applied the concept of entropy maximization from the field of physics to derive a family of spatial interaction models of which Ravenstein's gravity models are a specific instance (Wilson, 1967). The theoretical framework that Wilson used to derive his family of spatial interaction models is somewhat arbitrary and is far removed from any human aspect of the decision-making process.

A more behaviouristic basis for spatial interaction modelling became evident when it was realized that migration models were essentially similar in structure to the discrete choice models being applied to problems such as brand choice or mode of transport selection by economists (McFadden, 1978). This meant that spatial interaction models could be derived from principles of rational utility maximization, which are more intuitively applicable to migration destination choice than are statistical mechanics.

The latest evolution in spatial interaction modelling is logit-based discrete choice models that employ various means to represent the spatial structure of the destination choice set and how this affects spatial decision-making. This fourth generation of spatial interaction models is exemplified by the competing destinations model, developed by Fotheringham in the 1980s (Fotheringham, 1983, 1991) and the nested logit model adopted by geographers from econometricians and marketing analysts in the 1990s (Ben-Akiva, 1987; Evers, 1990). These models employ the concepts of destination accessibility and regional utility, respectively, to account for the impact of destinations' relative locations on how migrants cognize those destinations, particularly in terms of how migrants may group those destinations and therefore define a hierarchical choice set.

The premise that individuals arrange destinations into clusters for spatial decision-making purposes implies that they also store spatial information about those destinations in a similarly hierarchical manner. There is a great deal of evidence to suggest that place information is indeed stored in such a hierarchical mental representation, a variety of which is presented in chapter three, including: place information surfaces; counter-intuitive geographical 'facts'; positional and directional errors within and between destination clusters in recall and navigation studies; and, response times in place information recall experiments.

Employing a hierarchical mental representation of space that groups destinations into clusters brings into play another psychological phenomenon, the psychophysical theory that the perceived size of clusters (in general) is underestimated to an increasing degree as the cluster size increases. This phenomenon was investigated and has been extensively documented by Stevens (1975, 1978) and forms the basis for the inclusion of the accessibility statistic in the competing destinations spatial interaction model. The competing destinations migration model embraces the degree of destination clustering using an 'accessibility' variable.

The destination accessibility statistic represents the likelihood that a migrant will perceive a specific destination as being a member of a larger group of destinations, or in other words, it represents the likely size of the cluster to which a migrant will allocate a specific destination. In practical terms this means that, *ceteris paribus*, a destination with a high accessibility statistic is less likely to receive the individual attention from migrants that it merits in the destination selection process, because the number of destinations within a larger cluster will be underestimated compared to smaller clusters of destinations, and so a larger cluster will receive less attention as a whole than its actual size merits.

Another way of considering the accessibility of a destination is as the likelihood of that destination actually appearing in an individual migrant's choice set at all. According to Stevens' Psychophysical law, the higher an individual destination's accessibility value the more likely it will be overlooked within a larger destination cluster.

As well as impacting destination choice behaviour due to non-linear underestimation of destination cluster size, the spatial arrangement of potential migration destinations also impacts on the availability of destination information. It is reasonable to expect that the amount of information that a migrant has about a particular destination will be influenced by the number, size and proximity of that destination's neighbours. A more remote destination with fewer close neighbours is more likely, *ceteris paribus*, to be considered a 'place' in its own right and more of a migrant's information about such a place is likely to be specific to just that one place. In contrast, one would anticipate that more of a migrant's information about a destination with lots of large and nearby neighbours would be applicable to the cluster as a whole and not just to the specific destination – which would be of limited use when choosing between the migration destinations within that cluster.

This theoretical basis suggests that the accessibility variable of a competing destinations model of migration destination choice should have a negative parameter estimate, with higher values of the variable correlating with less frequent destination selection. This is indeed the pattern found in almost all applications of the competing destinations model (Atkins and Fotheringham, 1999; Fotheringham, 1986; Fotheringham, 1987).

The inclusion of an accessibility variable to incorporate the spatial competition effects into spatial interaction models is one of two methods investigated in this research. The second approach is to impose a spatial hierarchy on the migrant's choice set by allocating each destination to a specific cluster. The standard discrete choice logit model of destination choice is then extended to include an additional explanatory variable representing the utility of each destination cluster or region. This approach can be operationalised using the nested-logit model, which has been applied extensively in other non-spatial hierarchical choice applications, particularly in the fields of economics and marketing (Ben-Akiva, 1987; Evers, 1990).

There are a number of ways that the operation of the nested logit model can be understood, but the simplest is to consider it as two sequential logit calibrations. The first calibration is a traditional logit model of all the potential destinations from which a utility is generated for each destination. These utilities can then be summed within each destination cluster to give regional utility values that are incorporated as an additional destination description variable into a second logit model calibration. The results from the second calibration will be free of bias arising from the spatial structure of the destinations.

One would anticipate that destinations from regions with higher regional utility values would have a higher likelihood of selection *ceteris paribus* than those in regions with lower regional utility values. Thus, one would expect a positive parameter estimate is likely for the regional utility variable when the nested logit model is calibrated.

Perhaps the nested-logit model's most obvious shortcoming when applied to migration modelling is the imposition of an unrealistically rigid spatial hierarchy. It is highly unlikely that any one regionalisation could be an accurate reflection of the many varied cognitive spatial hierarchies of a large number of heterogeneous individual migrants. In order to address this issue probabilistic choice sets have been generated which in effect indicate the likelihood of each individual destination being grouped with each other destination. The nested logit framework has also been extended to produce a novel model called the weighted nested logit model that includes the utilities of probabilistically-defined rather than discrete regions as an additional explanatory variable in the second stage logit calibration.

Another criticism of the standard discrete regionalisation nested logit model is that the results of calibration are sensitive to the spatial organisation of the regionalisation used to define the choice sets. The weighted nested logit model also resolves this problem. Any two probabilistic regionalizations constructed from a sufficiently large number of separate discrete regionalizations will be very similar in structure and will therefore produce very similar parameter estimates when a weighted nested logit model is calibrated against them.

Novel techniques were developed and are presented in this thesis for the generation of the discrete and probabilistic regionalizations that are required in order to calibrate the discrete and weighted nested logit models of migration destination choice.

Discussion of epistemology

This research follows the traditional deterministic positivist approach to migration analysis that has been adopted by the majority of quantitative analyses of migration over recent decades. There has also been a good deal of qualitative research into this field. However, whilst this research has provided valuable insights into the many factors affecting the migration process, such approaches rarely provide insights that can be generalized and usefully applied in an aggregate-scale policy or planning context.

Whilst the aims of this research are primarily academic, and concerned with gaining a deeper understanding of the destination choice behaviour of migrants within Great Britain, it is also intended that the findings be influential in a wider context, by contributing to the next generation of migration models used by both public and private sector organisations for predicting future population distributions.

A further reason for adopting a data-driven, revealed preference approach to the analysis of migration destination choice rather than more opinion-survey-based methods, is that the behaviour under investigation may well not be entirely conscious and thus any interview or questionnaire-based analysis would be inadequate. Such analyses might provide valuable information as to the degree to which individual migrants are conscious of their destination choice behaviour, and of the mental representations of space that they construct to guide their destination choice behaviour. However, this research is concerned with demonstrating that the destination choice process underlying human migration is inherently hierarchical and potentially subconscious in nature, and in this context actions speak louder than words, so actions form the basis for the analysis.

It is interesting here to note that whilst the hierarchical migration models applied in this research, (presented in chapter 5), and the traditional migration models of the past, (presented in chapter 2) are superficially similar in operational structure they differ considerably in the principles guiding their derivation. Whereas the traditional models are derived from detailed examination of patterns in empirical data or the application of equally impersonal information theory, the hierarchical models are based on individual-level cognitive data structures and decision processes. This is a more humanistic approach to the derivation of aggregate migration models and addresses the traditional criticisms of spatial interaction modelling in general and migration modelling in particular, as having no solid personal basis. Positioning modern migration destination choice models within the framework of discrete choice modelling provides a very intuitive and personal rationale for the decisions under consideration, whilst also being scalable to facilitate meaningful application in traditional policy and commercial applications.

Another interesting trend in quantitative migration research exemplified by the analysis reported in this thesis is the increase in use of localized analysis. Computational advances and more disaggregated migration flow data now make it possible to calibrate models independently for each origin in a migration system. Results from such local, origin-specific calibrations demonstrate that a considerable degree of spatial variation in migration destination choice behaviour was previously being obscured by the averaging effect of global model calibrations (Fotheringham, 2000; Atkins, 1996). Such analysis starts to bridge the gap between the traditionally aggregate scale of quantitative migration analysis and the smaller individual/family/small group scale typically associated with qualitative migration studies.

Structure of thesis

This thesis is organised into four sections:

Section one introduces the research topic of hierarchical migration destination choice and places it within the wider context of migration research and cognitive psychology and establishes the specific goals of this research. The rationale behind the selected methodology is presented through a review of the past and current approaches to the analysis of migration destination choice. The often disparate fields of cognitive psychology and spatial information processing are then brought together to illustrate the development this research represents from traditional migration destination choice modelling and also to provide a theoretical justification for the selected methodology. The general operational approach adopted to achieve the stated research goals is then presented in terms of practical modelling tasks.

Section 2 describes in detail the methodology employed to achieve the research goals presented in section 1. This section of the thesis describes the various data used in the specification and calibration of the models, and the derivation of the migration system within which all analyses were performed. Also, the algorithms used to generate the hierarchical migration destination choice sets required for the application of the nested logit model are presented along with the derivation of the hierarchical destination choice models themselves.

Section 3 reviews the results obtained from the application of the various techniques described in section 2. The results of the traditional and hierarchical models are compared and the revealed migration destination choice behaviour patterns are interpreted. The remaining variation in observed migration behaviour not explained by the various calibrated models is examined and compared using various statistical and visualisation techniques.

Section 4 provides a summary of the key findings from this research and a concluding discussion of their relevance in a wider context. The known limitations of the research are identified and this feeds into suggestions for further research that could be conducted in this area.

Aims and Objectives

The research undertaken here reviews the current state of migration destination choice analysis and presents the existing theoretical and empirical evidence supporting the theory that the decision process underlying such destination choice is an inherently hierarchical process whereby migrants consciously or unconsciously cluster potential destinations and consider regional factors and spatial structure effects when choosing where to move. Further evidence in support of this hierarchical theory is then sought by calibrating a number of quantitative destination choice models derived from hierarchical principles, and comparing these results against those obtained from the traditional non-hierarchical model that has been the mainstay of migration destination choice analysis for many years. Comparisons are also made between the results of the various hierarchical models and also between migrant sub-groups disaggregated by gender, age and marital status, in order to further explore spatial and socio-demographic variations in migrants' destination choice behaviour.

The ultimate objective of this research is to make demographic modellers aware of the significant improvements that inherently hierarchical models provide over 'traditional' models of migration destination choice, such that these principles can be carried forward into more comprehensive and complex analyses of future population redistribution.

Chapter Two

Traditional Migration Analysis

Importance of migration

Patterns of internal migration and the redistribution of the population that they produce are of interest to many disparate groups in society, including: academics, public service planners, retailers, government departments and local authorities. This wide interest stems from the fact that migration is now the major factor in British population redistribution, with natural demographic change having become increasingly aspatial over the last century (Champion & Fielding, 1992). All these groups of people are concerned with future patterns of population distribution, be it to save lives by optimally locating ambulance stations, or to maximise profits through optimal store/depot positioning. An understanding of the process of migration is central to any attempts to predict future population shifts.

What is migration?

Intuitively this may seem a very simple question – answer: migration is changing one's address - and indeed the Office of Population Censuses and Surveys (OPCS)[1] definition of a migrant runs along these lines:-

> *"A 'migrant within one year preceding the Census' (often referred to simply as a 'migrant') is a person with a different usual address one year ago to that at Census."*
>
> *(OPCS, 1992)*

---

[1] The OPCS joined with the Central Statistical Office (CSO) in 1996 to become the Office of National Statistics (ONS).

However, when discussing the concept of migration, issues such as distance, duration and frequency of move should be considered (Plane & Rogerson, 1994). For instance, should students, who travel to and from university three times a year, sometimes over long distances, be considered migrants? Should the Smiths, who buy and move to the Jones' house next door, be considered migrants, despite the fact they have moved only 30 yards and have probably not altered their social network at all? Consider the seasonal worker who instructs skiers holidaying at Aviemore through the winter and then hires out donkeys on Blackpool beach all summer - should his twice yearly change of address be considered migrations? Even more fundamentally, should it be migrants or migrations that are counted?

In practice such considerations are often academic because the dataset being employed to investigate migration behaviour will dictate which aspects of migration can be studied, depending on how that dataset compiles and represents its migration data.

The availability and quality of data about internal migration in Great Britain has improved significantly from the situation in the late 1960s, which prompted Mackay to write:-

> *"The inadequacy of British migration statistics is almost legendary"*
>
> *(Mackay, 1969)*

Today, for the aggregate analysis of internal migration in Britain there are two very valuable, and very different, sources of data on migration: the OPCS Census of Population and the National Health Service Central Register (NHSCR).

Migration Data

Census Migration Data

Migration data derived from the 1991 Census of Population is based upon answers to the following two questions taken from the Census form:

> *7. Usual Address*
>
> *If the person usually lives here, please tick 'This Address'. If not, tick 'Elsewhere' and write in the person's usual address.*
>
> *For students and children away from home during term time, the home address should be taken as the usual address.*
>
> *For any person who lives away from home for part of the week, the home address should be taken as the usual address.*
>
> *Any person who is not a permanent member of the household should be asked what he or she considers to be his or her usual address.*
>
> *9. Usual address one year ago*
>
> *If the person's usual address one year ago (on the 21st April 1990) was the same as his or her current usual address (given in answer to question 7), please tick 'Same'. If not, tick 'Different' and write in the usual address one year ago.*
>
> *If everyone on the form has moved from the same address, please write the address in full for the first person and indicate with an arrow that this applies to the other people on the from.*
>
> *For a child born since the 21st April 1990, tick the 'Child under one' box.*
>
> *(OPCS, 1991)*

Such a derivation does not take account of multiple moves in the pre-censal year, which cannot be distinguished from a single move using the information from the Census form.

Thus, the Census counts migrants rather than migrations. This is a fundamental difference between the OPCS and NHSCR data.

The migrant information derived from Population Census returns is made available in a number of different formats. Numerous paper-based county and regional monitors of Census data are published, containing information on a wide range of topics, including migrants. A number of special topic volumes are also produced, including one on internal migration. The major concern of this thesis, however, is with machine readable data, in a form suitable for computation and visualisation. Increasing use of computers throughout the processing of the Census is producing increasing amounts of machine readable Census data at each Census. Data from the 1991 Census has been made available in the following four machine readable forms:

1. SAS/LBS – Small Area Statistics / Local Base Statistics
2. LS – Longitudinal Study
3. SAR – Samples of Anonymized Records
4. SMS – Special Migration Statistics

These are described below in more detail.

The standard Census tables, the Small Area Statistics (SAS) and the Local Base Statistics (LBS), include 100 and 96 tables, respectively, containing a wide array of data about the 113465 EDs and 10529 Census wards, respectively. Some of these tables relate to migration and contain a variety of information about migrants. OPCS (1991) provides a topic index to these tables together with a comprehensive listing of all SAS and LBS table layouts.

Longitudinally linked data about migrants is available from the Longitudinal Study (LS), based at City University, London, which links the 1971, 1981, 1991 and 2001 Population Censuses and vital registration data, using NHS numbers as the primary means of tracing individuals from one Census or registration to the next. This data is based on a 1% sample of the population (based on four specific birthdates), and because of this small sample size confidentiality concerns restrict the availability of data for smaller spatial units, restricting aggregate analysis with this dataset to larger spatial scales. This dataset has the advantage of having 10 year migration indicators for 1971-1981 and 1981-91, something not available from any other source (Atkins, 1995; Dale *et al*, 1993). A discussion of the potential benefits of longitudinal migration analyses over cross-sectional approaches, in the US context, can be found in Clark (1993).

A novel form of Population Census data, made available for the first time in Britain following the 1991 Census, are the Samples of Anonymized Records (SAR), which are effectively two samples of Census returns anonymized by removal of name and address data. There are two Samples, a 2% sample of individual returns and a 1% sample of household returns. The geographical coding of these Samples is not very detailed, with current addresses being coded to the 278 SAR areas, and migrant address one-year prior to the Census being coded only to the 11 standard regions so this dataset is also not suitable for meaningful migration destination choice analysis. This dataset is similar in nature to the Public Use Microdata Samples (PUMS), made available in the US following the 1990 Census. See Fotheringham & Pellegrini (1996) for a comparison of these two datasets, including examples of their application to migration analysis. The Census Microdata Unit (CMU) has produced a useful User Guide which contains further information on the SAR (CMU, 1993).

The SAR combines many characteristics of both macro- and micro-scale datasets, providing both national coverage, and high individual detail. Boyle's use of the SARs in his analysis of in-migrants to remote rural areas of Scotland and Wales demonstrates the potential of this dataset for the integration of macro- and micro-level methodologies (Boyle, 1995a, 1995b).

The only migration-specific datasets to be made available from the 1991 Census are two sets of migration matrices called the Special Migration Statistics (SMS) that disaggregate migration flows at several spatial scales by a variety of migrant characteristics. The SMS set 1 comprises ward-level matrices (10529 square) which disaggregate each migrant flow by broad age-group and gender. Set 2 is made up of district-level flow matrices (459 square) which disaggregate migrants by: age group, gender, marital status, ethnic group, limiting long-term illness, economic position, tenure and fluency in Gaelic or Welsh. The set 2 data provides a useful combination of spatial detail and migrant information - the district-level geography is sufficiently granular to permit aggregation to various functional district categorizations, such as OPCS District Types or Boyle's Migrant Profile Clusters (Boyle, 1993), whilst there is also sufficient variety of migrant characteristic information available to facilitate numerous analyses. Though the original release of these statistics was subject to some suppression (for reasons of confidentiality), the vast majority of the missing data has now been reliably imputed by researchers at the Centre for Computational Geography at Leeds University (Duke-Williams, 1995). A general review of migration trends from the 1991 Census SMS can be found in Atkins *et al* (1996). For further details about the 1991 SMS dataset see OPCS (1992) or Atkins (1995). Perhaps most importantly, the SMS is the only dataset of migration flows between areas, all others providing summary statistics about in-migration. This brings the researcher 'closer' to the process, and allows much more detailed analyses of what is undoubtedly a very complex process.

It is migration data from the SMS set 2 that is used to calibrate the various traditional and hierarchical migration models reported in this thesis. The district level geography of this dataset limits the volume of flow data to a level that can be analysed effectively using an inexpensive desktop personal computer.

NHSCR Migration Data

The most fundamental difference between the NHSCR and the Census migration data is that the NHSCR counts **migrations,** whilst the Census counts **migrants.** There are also major differences in the spatial detail and coverage of the two sources, which result from the different data collection methods.

The NHSCR migration data is derived from patient registrations with general practitioners. The NHS organises England and Wales and Scotland[2] into 115 Family Health Service Authorities (FHSAs), which equate to either shire counties, or single or combined metropolitan districts. If, following a change of address, a person registers with a new doctor, their move will be recorded in the NHSCR migration data. If however, they move within or between FHSAs but do not register with a new doctor, their move will not be included in the migration data. Because of this, the NHSCR data is susceptible to delays between migration and patient re-registration, which vary between age groups and gender according to health and health risk. However, on the plus side, NHSCR migration data is released every quarter, providing valuable temporal resolution.

Other Migration Datasets

Various datasets, including the labour force survey, Gallup polls and other minor surveys have also been used for migration research, but these are not of suitable sample size to facilitate reliable aggregate analysis. This is particularly true when attempting to undertake spatially hierarchical analysis as this requires sufficient spatial granularity of migration reporting areas that these can be clustered into a sufficient number of destination regions.

[2] Though no internal movement data is available for the 15 NHSCR areas in Scotland.

What is the question?

The analysis of migration has traditionally examined the process as two largely independent questions: (1) whether to migrate at all? And, if so, then (2) where to? Though both questions are crucial to the accurate prediction of future migration patterns, the research described here is concerned purely with the process of migration destination choice and does not address the challenge of modelling migration propensity.

As mentioned above, even within the topic of migration destination choice there is question as to whether it is migrant's moving to particular destinations that should be counted, or migrations to that destination. However, this is largely resolved by the dataset that is selected for the analysis – in this research the use of the SMS data from the Census of Population dictates that migrants be counted rather than migrations.

Qualitative migration analysis

Valuable insights into migration behaviour can result from small scale, qualitative analyses such as questionnaire surveys, often backed up by interviews with migrants. Such qualitative approaches can be very effective at uncovering the motivations and determinants influencing individual migrant's destination choice behaviour. Indeed, by the very personalized nature of such techniques they will often highlight factors that are likely to be missed by aggregate scale quantitative analysis of migration. However, there are two major drawbacks to small-scale qualitative studies of migration. First, the results are very difficult to generalise to a scale at which they are useful for policy formulation or strategic decision-making. And, second, they generally assume that all aspects of the migration destination choice process are conscious. This is intuitively unlikely to be the case – indeed Desbarats (1983) comments that:

"...further advances in behavioural prediction will depend on our willingness to question the widely held assumptions of utility maximization and volitional control and on our effectiveness in isolating those events that are beyond the control or consciousness of decision-makers but that nonetheless impinge upon their behaviour."

Desbarats (1983)

These are the main reasons that the majority of research into the process of migration destination choice has adopted a quantitative methodology.

There has been much debate over the relative merits of quantitative and qualitative approaches, which has created a widespread realisation that a research strategy integrating both types of analysis is required to develop in-depth understanding of migration. Articulating this view, in the context of the recent dominance of aggregate analysis, an Institute of British Geographers (IBG) limited life working party on migration suggested in 1991, that:

"While there remains much further work to be done at the aggregate level ... the Working Party believes that appreciable progress in the next few years is likely to derive from micro-level analysis ... and require researchers to have greater familiarity with ethnographic research methods."

(Champion & Stillwell, 1991)

This theme has been developed further by Halfacree and Boyle, who propose that:

"It is not enough to encourage greater cross-fertilisation between the varied strands of migration research ... Instead, we need to pay more attention to the very conceptualisation of migration itself"

(Halfacree & Boyle, 1993)

The recent availability of Population Census microdata offers some hope for the integration of both micro- and macro-level approaches to migration analysis.

In this context it is also worthwhile to consider that recent evolutions in spatial interaction modelling are derived from more humanistic principles based on individual's cognitive perception and representation of their environment rather than the purely empirical social physics of Ravenstein which underpinned the bulk of quantitative migration destination choice analysis until the mid 1970s. Indeed, a lack of widespread appreciation for recent developments in the field of spatial interaction modelling, particularly amongst qualitative analysts, has resulted in continued and often unjustified criticism of the field's updated epistemological basis (Sayer 1976, 1992).

Quantitative migration analysis

Within the field of quantitative migration destination choice analysis, studies range in their scope and complexity from a simple interpretation of descriptive statistics about net flows of migrants between regions, say, to the application of complex spatial interaction models or neural networks to simulate migration between functional groupings of enumeration districts.

Descriptive approach

Some degree of understanding of the migration process can be obtained by making intuitive inferences based upon summary descriptive information, such as plotting in and out migration characteristics against area's socio-economic characteristics. There are a great many studies of this type (Champion, 1999; Champion and Atkins 2000; Champion, 2001). However, such research often generates as many questions as it answers, as the unexpected layers of complexity are uncovered by the examination of summary data.

Neural spatial interaction models and genetic algorithms

A novel approach that has recently been applied to predicting spatial interaction is the use of neural networks. Neural networks, as their name suggests, draw their origins from the operation of neurons in the brain. A neural network is essentially an ordered group of processing nodes each of which takes one or more inputs, does some processing, and outputs a value. The early development and adoption of neural networks focused on their application to pattern recognition applications, but recent work that has drawn parallels between neural network analysis and traditional statistical methods has brought them wider acceptance and application (Cheng and Titterington, 1994; West, Brocket and Golden, 1997).

In a production-constrained migration modelling context a vector of destination characteristic values would be introduced to the network via a number of input nodes. These would typically forward on their output values to a hidden middle-tier which would forward on their output values to a single output-node which would output a vector of predicted flows to the various possible destinations. The primary challenge in such a neural network is determining the appropriate set of weightings that each node should apply to each of its various inputs to produce the best results. Initial applications of neural networks to spatial interaction in the early-1990s by Openshaw and Fischer optimized nodal weighting using gradient-based local minimization techniques (Openshaw, 1993; Fischer & Gopal, 1994). More recent adaptations and extensions make use of global search methods that can escape local minima to determine globally optimal weightings in a multi-model solution space (Fischer & Reismann, 1999; Fischer, Hlavackova-Schindler & Reismann, 1999, 2003; Fischer, 2002).

Another challenge when specifying neural spatial interaction models is determining the appropriate network topology that is best suited to a particular spatial interaction problem. A novel development in this area has been the application of genetic algorithms to iteratively identify the optimal topology. Genetic algorithms are based on the concepts of evolutionary biology insofar as they maintain a pool of possible solutions each of which has a score or *fitness* associated with it (Holland, 1975). In an iterative fashion the fittest of the stored solutions will be combined to produce new solutions that will be tested and awarded fitness scores. The better performing of these will displace lesser solutions from the finite solution set – or in other words the fittest solutions will survive in the solution *gene pool.* Over time the solutions that remain will be more and more optimal. Fischer and Leung applied such a evolutionary genetic approach to determining the optimal neural network topology with which to predict telecommunications traffic within Austria, (Fischer and Leung, 1998). This can be computationally intensive as, depending on the complexity of the problem being solved, a very large number of iterations may be required before an apparently optimal solution is discovered, (Turton, Openshaw and Diplock, 1997).

Neural networks that have been sufficiently well 'trained' using historical data can often provide comparable or better predictions of future spatial interaction trends than non-neural spatial interaction models (Wier & Phoha, 1999; Fischer, Hlavackova-Schindler & Reismann, 1999). However, they not as straightforward to interpret in a meaningful way as a parameterized logit model, or as Reggiani *et al* put it, the neural spatial interaction approach *"is less easily interpretable from social science motives"* (Reggiani, Nijkamp and Tsang, 1997). Thus, they are a less useful means of deepening understanding of the psychological decision-making processes underlying the spatial interaction patterns that they are predicting. For this reason, they are not considered further in the current research.

In order to investigate the causality of migration destination choice it is preferable to calibrate parametric models of the migration process which more directly show the effect upon migration behaviour of a variety of potential explanatory variables, whilst controlling for variation in all other variables.

Heuristic-based simulation

Another approach to migration analysis is to aggregate the simulated behaviour of individuals based on simple motivational heuristics. Most examples of this approach greatly over-simplify the motivating factors and have typically been applied in the fields of ecology and historical anthropology where it is the pattern of spread of an animal or civilization that is the subject of study (Young, 2002). Though, simulations of semi-random rural-urban human migration have been used to demonstrate commonalities in the urban structures that result from the aggregation of *'innumerable unpredictable events'* (Wong & Fotheringham, 1990).

Migration modelling

The hierarchical models of migration destination choice presented in this thesis represent a significant evolution of migration destination choice modelling. Any such methodology development should be considered in relation to the historical context of the field. In the case of migration modelling, the historical record is often misrepresented or at least misunderstood, so a review is presented here.

A good deal of the misperception of spatial interaction modelling in general, and migration destination choice modelling in particular, as outmoded and non-humanist analytical tools arises from the fact that modern migration models appear structurally very similar to those developed over a century ago. In reality, however, the derivation of migration models has evolved a great deal, particularly in the last four decades. Four developmental phases have been identified in the history of spatial interaction modelling, (Fotheringham, 2000). Whilst there is some inevitable overlap in their chronology, each phase marks a fundamental advance from its predecessor. The phases identified by Fotheringham are:

1. Social physics (mid1800s-late1960s)
2. Statistical mechanics (late1960s-early1980s)
3. Aspatial information processing (early1980s-early1990s)
4. Spatial information processing (early 1990s onwards)

1. Social physics

The earliest published records of serious academic research into human spatial interaction date back to the middle of the nineteenth century when first Carey and then Ravenstein examined the factors influencing the movement of people between settlements (Carey, 1858; Ravenstein, 1885). Ravenstein observed that migration from one area to another was greater if the areas were located closer together and if they were larger. Coming from a background in physics, Ravenstein recognized the parallels between this phenomenon and the gravitation pull which every body of mass exerts on all others around it. Because of his use of this analogy, Ravenstein's field of research became known as 'social physics' and his approach to the analysis of spatial interaction was termed gravity modelling. Ravenstein's original gravity model formulation was very simplistic and is presented in equation 2.1 below.

$$T_{ij} = k\frac{P_i P_j}{d_{ij}} \qquad \textit{(Eq.2.1)}$$

Where:
T_{ij} is the spatial interaction (traffic) from origin i to destination j
K is a balancing factor
P_i is the population of area i
P_j is the population of area j
d_{ij} is the separation of areas i and j

Equation 2.1: Original basic form of Ravenstein's gravity model of spatial interaction.

In equation 2.1 T_{ij} represents the number of migrants moving from origin i who choose to move to destination j; P_i and P_j represent the population of origin i and destination j, respectively; and d_{ij} represents the separation of areas i and j.

Ravenstein's model is based on the observation that population is the most important 'mass variable' contributing to a migration destination's 'gravitational pull' over migrants, something which is still generally accepted to this day. The first evolution of Ravenstein's original gravity model was the introduction of additional explanatory variables describing the potential destinations. In keeping with the social physics metaphor, these explanatory variables were often referred to as mass variables. Another development was the use of exponents on the various explanatory variables in recognition of the fact that the same explanatory variable might have a different degree of influence in different situations and to different groups of migrants. The formulation of a more general gravity model is presented in equation 2.2 below.

$$T_{ij} = k\frac{O_{i1}^{\alpha 1}\, O_{i2}^{\alpha 2} \ldots O_{if}^{\alpha f}\, D_{j1}^{\lambda 1}\, D_{j2}^{\lambda 2} \ldots D_{jg}^{\lambda g}}{d_{ij}^{\beta}} \qquad \textit{(Eq.2.2)}$$

Where:

T_{ij} is the spatial interaction (traffic) from origin i to destination j
k is a balancing factor
$O_{i1}^{\alpha 1}$ is parameterised explanatory variable 1 describing origin i
$D_{j1}^{\lambda 1}$ is parameterised explanatory variable 1 describing destination j
d_{ij}^{β} is the parameterised origin-destination separation variable

Equation 2.2: General form of the gravity model of spatial interaction.

In equation 2.2, a vector O_i of f explanatory variables describe the origin i and the vector D_j of g explanatory variables describes destination j. The model calibration process estimates for the parameters $\alpha 1 \ldots \alpha f$, $\lambda 1 \ldots \lambda g$ and β, which can then be interpreted to determine the nature and strength of effect of each explanatory variable.

Gravity modelling, as popularized by Ravenstein, held sway for a considerable period of time. One of the reasons for the lack of further development in this area was the lack of processing power available to automate the calibration of the models. With the advent of the microprocessor, increasing availability of computing power in the 1960s revitalized interest in spatial interaction models as it became faster to calibrate more and more complex models.

At around the same time tensions were arising within the human geography research community between some practitioners of qualitative and quantitative analysis. It was considered a major weakness of Ravenstein's 'laws of migration' and of research based upon them, that they were essentially derived from examination of recorded migration patterns, with no real theoretical basis underpinning them. In the late 1960s Wilson presented a theoretical framework for spatial interaction modelling based on maximizing the entropy of a migration system, an idea borrowed from the field of statistical mechanics.

2. Statistical mechanics

Wilson considered the problem of spatial interaction modelling from a different perspective (Wilson, 1967). He considered migration in terms of 'macrostates' and 'microstates', where a macrostate describes total migration flows between areas and a microstate lists which individual migrants moved between each origin-destination pair. Most macrostates can result from a number of microstates, and Wilson's central premise is that *ceteris paribus* the macrostate that can be arrived at via the largest number of microstates is the macrostate that is most likely to occur.

Consider the example (after Fotheringham 2000) of a simple migration system containing just three places, A, B and C in which five migrants leave place A. There are six macrostates that are consistent with this information. Table 2.1 below tabulates these with the corresponding number of possible microstates.

Macrostate	M_{AB}	M_{AC}	Number of microstates
1	0	5	1
2	1	4	5
3	2	3	10
4	3	2	10
5	4	1	5
6	5	0	1

Table 2.1: Wilson's macro- and micro-states.

A gross out-migration population of five migrants offers 10 distinct migrant pairs that could be the specific individuals moving from place A to place B in macrostate 3 above, and each of those 10 migrant pairs is a microstate. In this example, Wilson would conclude that macrostates 3 and 4 are *ceteris paribus* most likely to occur because there are more possible microstates that can lead to that outcome.

By making assumptions about adequate flow sizes Wilson was able to demonstrate mathematically that finding the macrostate that could result from the maximum number of microstates was equivalent to maximizing the entropy of the distribution of flows within the migration system. This led to his approach to spatial interaction modelling being referred to as entropy modelling. Wilson further defined various constraints which can be imposed on the maximization process giving rise to a whole family of models. One of these models is identical in form to the simple gravity model presented above. Another, which has become the most commonly used model in migration analysis, is the production-constrained spatial interaction model, where predicted gross out-flow from all areas is constrained to match total observed gross outflow.

Whilst it could be argued that the ultimate results of Wilson's work was to arrive at very similar models to those which had previously been formulated, it is important to recognize that his work provided a theoretical basis for the derivation of the models, which was more acceptable to the research community of the day than the broadly empirical basis on which Ravenstein's gravity models had been postulated.

There are, however, criticisms which can be levelled against Wilson's modelling framework. Most notably, his derivation makes use of questionable assumptions of significant flow sizes which will not be met in many real world spatial interaction situations, including migration analysis. Also, to many human geographers a model based on maximizing the entropy of a statistical distribution is too far removed from the individual-level decision-making that determines migration to hold much credibility. The next major development in spatial interaction modelling would address this latter issue directly, by deriving destination choice models based on the very rational principal of utility maximization.

3. Aspatial Information Processing

A major step forward was made in the early 1980s when it was realized that spatial interaction models could also be considered as discrete choice models based on the logit formulation, i.e. a binary regression with continuous explanatory variables and a Gumble distributed error term. This provided a derivation of the spatial interaction model based on individual-level information processing and decision-making behaviour (Boots, 1988; Wrigley, 1988; Maier and Weiss, 1990).

Discrete choice models apply the principle of rational utility maximization to model an individual's choice between mutually exclusive values of a categorical variable. Before the early 1980s this type of model had only been widely applied to aspatial choice sets such as different brands of chocolate or means of commuting. It was a beneficial development in the field of spatial interaction modelling when it was realized that such models could also be usefully applied to spatial choice sets, such as migration, shopping and vacation destinations.

The logit model predicts which potential alternative will be selected from a choice set by assuming rational utility-maximizing behaviour by the decision-maker. It is assumed that a rational decision-maker will select the specific alternative that offers the highest utility. However, because of imperfections in data and the unknown effects of missing information about alternatives, the discrete choice model is usually formulated as a probability of a specific decision-maker selecting any particular alternative. This formulation of the logit model, presented below in equation 2.3, calculates the probability of an alternative being selected as the ratio of that alternative's utility to the sum of the utilities of all potential alternatives.

$$P_{ij} = exp\,(V_{ij}) \,/ \sum_{K} exp\,(V_{ik}) \qquad (Eq.2.3)$$

Where:

P_{ij} *is the likelihood of a migrant from origin i selecting destination j*
V_{ij} *is a vector of variables describing destination j for a migrant from origin i*
V_{ik} *is a vector of variables describing destination k for a migrant from origin i*

Equation 2.3: Basic logit model formulation

When considering the behaviour of a population of homogeneous individuals as a whole, the probability of an individual from that population selecting any specific alternative can be factored by the size of the population to produce a 'share model' which predicts how many individuals from that population that will select each alternative. The logit formulation presented in equation 2.4 below is 'production-constrained', meaning the population under consideration is defined as all migrants leaving origin i (represented by the M_i term in equation 2.4).

The utility variable V_{ij} can also be expanded to show the various variables that make it up. When applied to migration destination choice modelling these variables will be the same origin and destination characteristic variables that would be used in the more advanced gravity model formulation shown above in equation 2.2. The expanded share formulation of the logit model is presented below in equation 2.4.

$$M_{ij} = M_i \frac{exp\,(O_{i1}^{\alpha 1}\, O_{i2}^{\alpha 2} \ldots O_{if}^{\alpha f}\, D_{j1}^{\lambda 1}\, D_{j2}^{\lambda 2} \ldots D_{jg}^{\lambda g})}{\sum_{k} exp\,(O_{i1}^{\alpha 1}\, O_{i2}^{\alpha 2} \ldots O_{if}^{\alpha f}\, D_{k1}^{\lambda 1}\, D_{k2}^{\lambda 2} \ldots D_{kg}^{\lambda g})} \qquad (Eq.2.4)$$

Where:

M_{ij} *is the migration from origin i to destination j*

M_i is the total out-migration from origin i
$O_{il}^{\alpha l}$ is parameterised explanatory variable 1 describing origin i
$D_{jl}^{\lambda l}$ is parameterised explanatory variable 1 describing destination j

Equation 2.4: Full formulation of logit model of migration destination choice

Of course, real world decision makers are not homogeneous individuals. Whilst no aggregate discrete choice model can explicitly cater for each individual's particular decision making criteria, the formulation of the logit model does include a random component of each destination's utility which essentially accounts for:

- individual variation in the contribution of particular alternative attributes;
- individual variation in the perceived values of the included alternative attributes;

Assuming different distributions for this 'random' component in such utility maximizing models gives rise to a family of discrete choice models. The most computationally tractable of these models is the logit model which assumes a Gumbel (or type I extreme value) distribution. The only other commonly used model in this family is the probit model, which assumes a normal distribution. The logit model was applied here partly because it is less mathematically complex and computationally intensive, and also because, to quote Ben-Akiva and Lerman: *"there is still no evidence to suggest in which situations the greater generality of multinomial probit is worth the additional computational problems resulting from its use"* (Ben-Akiva and Lerman, 1987).

The model described in equation 2.4 is hereafter referred to as the traditional migration destination choice model. It is this model against which the performance of the hierarchical models, derived and discussed below, will be compared.

4. Spatial Information Processing

The fourth and current generation of spatial interaction models that have been applied to the analysis of migration destination choice extend the third generation aspatial models by accounting for the spatial structure in the choice set. Failure to consider the choice set's spatial characteristics leads to several problems which are described in the next section. A number of fourth generation migration destination choice models are presented in detail in chapter 5.

The drawbacks of traditional migration modelling

Fotheringham (1983, 1986, 1991) and others have pointed out that the traditional logit approach to modelling of discrete choice has a number of important characteristics which, whilst not problematic in the model's more traditional application to modelling choice between aspatial alternatives, are of concern when location is a characteristic of the alternatives in the choice set. These undesirable properties of the logit discrete choice model are:

- the Independence from Irrelevant Alternatives (IIA) property
- regularity
- scale of processing

The IIA property of the logit model means that the relative popularity of any two particular alternatives is unaffected by the introduction of a third alternative. It is intuitive that this is not appropriate when modelling a spatial choice process, such as the selection of a migration destination. In such a case an additional alternative is unlikely to have equal effect on both destinations as it is unlikely to be located exactly equidistant between them. The attractiveness (or 'repulsiveness') of adjacent areas is highly likely to affect in-migration into any particular destination.

Exactly how the proximity of attractive or unattractive neighbours will affect the popularity of a particular destination is not so self-evident. It is entirely conceivable that a situation could arise whereby the popularity of an area could be increased by the addition of more or less appealing alternatives close by, either due to the agglomeration effect of a popular neighbour, or the reduced competition from an unpopular neighbour. However, a logit formulation of destination choice does not permit this to occur. A logit model of destination choice will only ever predict the same or less migration to any specific destination after the addition of a new potential destination into the choice set. This restrictive property of the logit model of migration destination choice was termed *regularity* by Huber *et al* (1982).

Both the IIA and Regularity properties above are inappropriate to the modelling of spatial choice, and models based on these assumptions will inevitably be biased and produce inaccurate destination choice predictions as they cannot make any allowance for the spatial structure of the choice set under consideration.

Another assumption of the aspatial logit model is that each potential destination receives equal attention from and consideration from each migrant. Both intuition and various studies in the field of cognitive psychology suggest that this is not likely to be an accurate representation of reality. The following chapter will discuss spatial information storage and processing in order to justify a hierarchical approach to migration destination choice modelling.

In response to these shortcomings the 1980s saw an increased consideration of the effects of human spatial cognition upon the migration process: how a migrant's perception of his or her environment affects the way that they select a migration destination. A more humanistic approach was adopted to the derivation of migration models that are based on individuals' decision-making processes and which account for the effects of spatial structure in the destination choice set. This new approach overcame the undesirable properties of the traditional logit model discussed above. It also signalled some degree of convergence of ideas from the qualitative and quantitative approaches to migration research as a more humanistic consideration of the migration destination choice process started to permeate the thinking of traditionally 'aggregate analysts'.

Summary

This chapter introduces the key topics in migration research: what is a migrant? What migration data are available? Whether to count migrants or migrations? The variety of analytical approaches to migration is highlighted with particular emphasis on the evolution of migration modelling through a number of well defined stages:

1) social physics: Ravenstein's gravity modelling
2) statistical mechanics: Wilson's entropy maximizing models
3) aspatial discrete choice: logit modelling
4) spatial discrete choice: logit modelling

The limitations and drawbacks of these traditional approaches are presented as justification for the development and application of a next generation of migration models that are derived from hierarchical principles and incorporate the spatial structure of the destination choice set.

The next chapter presents the theoretical basis for this new generation of hierarchical migration destination choice models.

Chapter Three

Theoretical basis for hierarchical destination choice

Introduction

Chapter one introduced what was meant by flat and hierarchical migration destination choice, and chapter two went on to discuss further traditional non-hierarchical approaches to the modelling of migration. This chapter examines the theoretical basis for the proposition that migration destination choice is a hierarchical process and then describes the theoretical underpinnings of a number of approaches to incorporating a spatial hierarchy into migration models.

There are a number of arguments against a simultaneous flat-processing theory of spatial choice behaviour in general, and of the migration destination choice process in particular. These objections fall into the following categories:

- Availability of information about choice alternatives;
- Cognitive storage and processing of spatial information;
- Empirical results of hierarchical choice modelling.

This evidence is presented in detail in the following sections.

Availability of destination information

It is highly unlikely that each migrant will even be aware of the existence of every potential destination let alone have sufficient information about them to be able to perform the comparisons necessary to make an optimal destination choice. It can be argued that this lack of information about some potential destinations does not invalidate the traditional migration modelling framework as such variation in destination information is partially modelled by the distance explanatory variable in the traditional model. However, it is surely misleading to calibrate parameters of a migration model based upon the effects of destinations which migrants are not even aware of? Because the distance variable represents both the likely destination information as well as the economic cost of moving, this parameter could perhaps be calibrated usefully across all destinations, but it seems erroneous to allow the house prices or unemployment rate of unknown destinations to affect the parameter estimates for these variables as they will if the model is calibrated across all destinations.

It is better to separately model the destination choice set composition for migrants from each origin, and then to calibrate a migration destination choice model across the identified subset of destinations. As will be seen later, this is in effect what the competing destinations and nested logit achieve by including an additional explanatory variable representing the likelihood of a destination actually being in a migrant's choice set. Although distance and population are almost certainly the key determinants of the level of destination information, it is unconvincing to argue that the distance and population variables in a traditional migration model already take account of choice set definition, as these variables also act as destination attractiveness variables, and it is impossible to deduce from the distance and population parameter estimates the relative effects of choice set participation and economic cost of moving, which *ceteris paribus* reduce and increase respectively with distance.

It is in considerations such as these that a humanistic derivation can seem at odds with the aggregate scale of application of the model. For although any individual migrant is highly unlikely to compare all possible destinations, it is not unreasonable for each and every destination to be considered by at least one migrant amongst all those who leave a particular Local Authority District. If there are several thousand migrants leaving a district it is much more likely that each possible district will be within the choice set of at least one migrant. Fortunately, the additional variable in the competing destinations and nested logit models of migration destination choice account for this by representing the likelihood of a particular destination appearing in the choice set of migrants from a particular origin.

As well as determining the likelihood of being considered by any specific migrant, a destination's population and distance from an origin, amongst other variables, will most likely also have an effect upon the quality of the information which migrants have about that destination. However, it is very hard to do anything more than hypothesise about such a qualitative issue, as reliable data do not exist and would be very difficult to gather.

With regard to the quality and nature of spatial information that migrants have about destinations, a distinction can be drawn between knowledge gained about those potential destinations through personal experience versus that gathered second-hand through description from maps, friends, family and the media (Presson and Somerville, 1985). It is not unreasonable to expect that the source will also likely have an effect upon the quality of destination information. Indeed, Thorndyke and Hayes-Roth (1982) demonstrated that mental spatial representations acquired from study of maps are more accurate than those derived from personal experience in an environment. However, Thorndyke (1981) also demonstrated that the performance benefits associated with map-learning are matched by sufficient experience in the environment. Similarly, Garling, Böök and Ergezen (1982) showed that representations acquired from primary experience became more 'map-like' with experience, as the various routes that an individual travels and learns overlap and become more interconnected. This would facilitate more accurate direction and separation estimation such as would be made with the use of a map.

Such studies also raise the question of how the quality of spatial information is measured. To get accurate distances one might assume careful measurement of roads or straight lines on a map would be ideal, but in human terms the transit-time distances that would more likely crop up in conversation are often a more useful measure of distance, as 100 miles along twisting B-roads is generally not comparable with 100 miles down the M1.

Similarly population size is a much better predictor of the information that a migrant may have about a place if it is considered within the context of its surrounding areas. For instance, Norwich is a small city and Penzance a small town, but because there are no other large settlements very close to them they are probably better known and distinctly characterized by migrants than similar sized areas in the outer London boroughs or on the outskirts of other major conurbations.

Of course, all other things being equal, information about more distant areas is still likely to be more Spartan, however, since a migrant is less likely to know people from whom to get information about more distant places and they are more likely to have only large-scale atlas maps of such less local areas.

Another key aspect of destination information availability is the extent to which migrants are capable of effectively storing and processing the spatial information that arises out of their personal experiences or reading or discussions with others. The following section reviews research in the field of cognitive psychology and examines this issue further.

Cognitive representation of spatial information

The field of cognitive psychology is concerned with the mechanisms of perception, storage and processing of all types of information. It has been recognized for some time that an understanding of the mental representation of spatial information is essential to gaining meaningful insights into human spatial behaviour (Golledge and Zannaras, 1973; Golledge, 1977; Cadwallader, 1979). Thus, the internal representation of spatial relationships has received much attention within this discipline. An understanding of migrants' internal representation of spatial information is useful when developing spatial interaction models to accurately predict individuals' migration destination choice behaviour.

It is only more recently that a number of studies have suggested that such spatial information is mentally represented, at least partly, in a hierarchical form. The majority of these studies draw conclusions from the ability of experimental subjects to judge distance and directional relationships between places.

1. Directional estimation errors

Another interesting distinction that has been demonstrated between spatial information obtained from personal experience and that obtained by secondary means, is that the former tends to be less orientation specific. Locations and routes learned first-hand appear to be coded relative to each other but not in relation to any larger frame of reference, whereas information from maps is often stored relative to the natural orientation of the map, usually North-South, (Evans and Pezdek, 1980; Sholl, 1987).

A number of studies have demonstrated that people tend to apply rules, or heuristics, to spatial data to make them easier to visualize, remember or explain to others. Tversky (1981) demonstrated experimentally the existence of two spatial heuristics: rotation and alignment which individuals bring to bear upon directional information. The effect of the *rotation* heuristic is to cause similarly (but not identically) oriented axes in a figure to be considered as being oriented in the same direction as a means of simplifying the directional information. The *alignment* heuristic results in similarly aligned and proximal axes to be aggregated together. Obviously, both of these heuristics serve to reduce the overall amount of spatial information in a system. Tversky also demonstrates that these simplifying heuristics *'may be adopted in storage, where spatial positions are difficult to encode, as well as in inference, to fill gaps of knowledge'* (Tversky, 1981). This demonstrated tendency to rotate and aggregate similar axes suggests that a single instance of the directional data is being encoded to memory for the group of axes, or in other words spatial information is being stored hierarchically.

Directional estimation error was also the subject of a study by Stevens and Coupe (1978) which asked subjects to plot the relative orientation of a number of pairs of places, including: Toronto and Portland, Oregon; Montreal and Seattle, and; the Atlantic and Pacific entrances to the Panama Canal. All three of these example pairs have counter-intuitive orientations: Toronto is south of Portland, Oregon; Montreal is south of Seattle, and; the Atlantic entrance to the Panama Canal is West of the Pacific entrance. However, Stevens and Coupe's experiment showed that the majority of subjects oriented these three pairs of places incorrectly. This suggests that the subject's orientation of the place pairs may have been based on relative positional information stored at a larger geographical scale in a hierarchy of spatial information, rather than being based on the locations of the places themselves, i.e. the subject knows that Canada is north of the US and therefore assumes that Toronto is north of Seattle.

Stevens and Coupe (1978) also performed experiments based on artificial spatial systems in which labelled points were positioned within regions. These spatial systems were presented to subjects for a short period of time and then removed and questions were posed about relative locations of points in the system. Directional estimation errors were found to be significantly higher when subjects were estimating orientation between pairs of points that matches the regional orientation. This suggests that hierarchical encoding of location information is used to simplify and optimize storage of spatial information from arbitrary images as well as for geographical configuration information that may be learnt by other means than images or maps.

McNamara (1986) performed similar experiments and made similar observations of between-cluster orientation judgements being affected by the relative positions of the containing regions.

Mark (1992) surveyed many subjects' perceptions of the latitude and longitude of a number of places, all relative to the city of Buffalo in New York state. His findings confirmed Tversky's rotation heuristic, with subjects' estimates of longitude relative to Buffalo, NY suggesting a North-South 'isoline' that was rotated to match closely the orientation of the East Coast of the US. These findings considered alongside those of Tversky and of Stevens & Coupe led Mark to conclude that their results showed *'systematic distortions of geographic configurations that are consistent with hierarchical knowledge representation'*.

Within the context of Great Britain it is interesting, and counter-intuitive to note that Bristol is in fact East of Edinburgh, and also that London is closer to the North Pole than Toronto. The former fact can be attributed to the some positional information about places being inferred on the basis of which coast of the country they are on. The London-Toronto 'fact' appears counter-intuitive largely because of climatic perceptions. It is tempting to assume that because Canada is colder and more snow covered than the UK, that it must therefore be further north, when in fact more complex meteorological issues are at play to confound basing latitudinal assumptions upon the weather.

Kuipers, Tecuci and Stankiewicz have proposed a theory supported by experimental and empirical data, suggesting that direction estimation and route-finding is based on a hierarchical route map, made up of a core 'skeleton' of known routes, with more fine-grained local route information being used at the origin and destination (Kuipers, Tecuci & Stankiewicz, 2003). Whilst the local route knowledge will inevitably vary considerably between individuals based on personal experience and geographical history, it is not unreasonable to expect more commonality in the core 'skeletal' route knowledge that will be shared by many individuals.

2. Distance estimation errors

Bias in distance estimation has been shown by Cadwallader (1979) to vary depending upon both the methodology employed and the spatial scale at which experiments are performed. This scale effect has been considered by many researchers, many of whom draw a distinction between small and large spatial scale. These definitions vary but in general 'small' in this context means up to the size of a city, and large typically refers to a large region or an entire country. In such terminology the research undertaken here relates to large scale spatial cognition (Curtis and Fotheringham, 1995).

Whilst distance estimation bias poses no inherent methodological problem for the purposes of the current research, it is interesting to consider because a number of experimental studies have demonstrated patterns in distance estimation error which suggest a hierarchical method of encoding and processing spatial information.

Hirtle and Jonides (1985) and McNamara (1986) showed that within-region distances estimates are more often under-estimated whilst between-region distance estimates are usually over-estimated. Experiments also demonstrate that natural region-forming boundaries, such as rivers, major roads and railway lines, can induce similar distance estimate error, with distances crossing such boundaries often over-estimated, and the distance between places not separated by such a landscape feature usually being under-estimated (Canter and Tagg, 1975; Newcombe and Liben, 1982).

Various studies have concluded there is a distance bias in spatial knowledge gained from personal experience by demonstrating that distances are usually over-estimated for more 'complex routes' which have more turns or landmarks (McNamara, Ratcliffe and McKoon, 1984; Briggs, 1973; Allen, 1981; Byrne, 1979; Sadella and Magel, 1980; Sadella and Staplin, 1980).

Sandberg, Huttenlocher and Newcombe (1996) demonstrated that location information, whether encoded as X and Y coordinates or as angle and distance, is stored both as categorical and as a continuous representation relative to some frame of reference. This is a crude form of hierarchical representation, the top level of which shows which locations are, say, North East of a reference point, with deeper access to the next level of the hierarchical representation being required to determine how far North East those locations are.

It is also interesting to note that, as demonstrated by MacEachren (1992), errors in distance estimation are more dependent upon the nature of the spatial task being performed (i.e. on information decoding) than are direction estimation errors which are broadly consistent regardless of the task at hand implying that the information decoding process has little impact on the accuracy of direction estimates. This could arise due to directional information being somewhat visual in its storage mechanism (i.e. the category North East and a specific direction can be associated with a mental image of a vector at a certain angle to the vertical). Distance information, on the other hand, is more likely to be categorized as a categorical component (A is east of B) along with a quantitative component (A is 56 miles east of B). The quantitative element of distance information is more likely to vary between individuals given variation in peoples' ability to accurately assess distance. Furthermore, it is more likely that some distance estimates will be inferred from a combination of categorical and quantitative information, whilst others will require accessing of quantitative information only. Different spatial tasks might give rise to different ways of estimating distances, which could explain why errors in distance estimates are more variable between spatial tasks than are directional estimates.

These findings are echoed by Kitchin (1997) who demonstrated experimentally that spatial processing approach varies with the task at hand as well as based on the underlying mental representation of the spatial information. Sherman, Croxton and Giovanatto (1979) found similar results when comparing different tests of individuals' distance estimation ability.

3. Speed of spatial cognition

Speed of confirmation/contradiction of stated geographical facts also provides evidence that hierarchical storage of spatial information underlies such processing. For instance, one might expect that response times to the statement 'Nottingham is south of Edinburgh' would on average be faster than to the statement 'Nottingham is south of Newcastle', because Edinburgh is in Scotland and Nottingham in England, so spatial information from a higher country-level in a hierarchical mental representation can confirm that Nottingham is indeed south of Edinburgh. However, comparison of the locations of Newcastle and Nottingham requires a more time-consuming drill down to lower levels of the hierarchy. This theory is supported by the work of Maki (1981) who confirmed faster distance estimation times for inter-regional than for intra-regional distances, and also of Wilton (1979) who demonstrated similar results for between-cluster and within-cluster orientation judgements. Furthermore, it suggests that a stored mental spatial hierarchy is only minimally queried, and/or that when building up mental representations of space we are occasionally 'lazy' and store information at a more general level than is always appropriate.

It is possible that such spatial relationships between places, which have to be generated from the relationships between 'higher-level' place groupings, could themselves be stored once they are calculated. The question might then be whether this newly calculated relationship is stored in the main long-term spatial representation, or as an isolated short-term fact? The latter is arguably more likely, as there is general consensus in the literature that human memory organisation is optimised for storage capacity rather than access speed, and redundant storage of spatial relations would tend to contradict this general design characteristic (Kosslyn, 1984; Wood, 1983; Squire, 1987; Farah, 1988).

4. Aspatial information storage and processing

Clearly there is much experimental evidence suggesting that spatial information is mentally stored in a hierarchical structure. There is little consensus, however, concerning the mental representation of the aspatial information associated with spatial information. For instance, is the fact that a building is a library or houses a Geography department associated with the spatial information about that building's location?

Kosslyn (1987) proposes that when learning a spatial layout two independent memory structures are constructed, a categorical memory structure, containing information such as relative locations and adjacency, and a coordinate spatial structure which stores information about absolute spatial location. It is proposed that only the former of these structures is hierarchical with the latter adhering more to the traditional 'metric' memory model. Interestingly, they further propose that these structures are physically located in and processed by opposite lobes of the brain. Their evidence for this theory is based upon the results of experiments concerning object recognition and image generation and whilst many of the experimental assumptions have been verified in the visual domain, they have not been rigorously explored in the context of spatial memory.

5. Other studies

Curtis and Fotheringham (1995) build on the work of Gould and White (1974) regarding preference surfaces to construct recall surfaces and to model the determinants of place recall. As one might expect the likelihood of a place being recalled was found to be positively related to a place's population and also to whether or not it was categorized as a state capital. It was found to be negatively impacted by a place's distance from the subject and also by the extent of spatial competition that a place experiences due to the relative locations of its neighbours. Curtis and Fotheringham interpreted these results as evidence of a hierarchical mental representation of spatial information. They reasoned that a hierarchical representation would group places into clusters and that according to Psychophysical principles proposed by Stevens (1975) the membership of larger clusters would be underestimated. This would lead to a place with more neighbours being recalled less often *ceteris paribus* than a place that is more remote from its neighbours. The parameter estimates from Curtis and Fotheringham's recall model calibrations confirm this theory.

A hierarchical theory of the mental representation of spatial information is also appealing as it has direct parallels with the hierarchical structures which are generally accepted as underlying some aspects of linguistic memory (Cienki, 1989; Herskovitz, 1986; Talmy, 1983). It is intuitive to assume that the nature of our language when talking about places, locations, directions...etc. is directly influenced by the way in which the spatial information being described in mentally represented.

Empirical evidence

Another source of evidence in support of a hierarchical theory of spatial information storage and processing is the results from existing research applying hierarchical choice models. In migration research, Fotheringham (1987, 1991), Curtis (1991) and Atkins & Fotheringham (1999) have demonstrated that the application of implicitly hierarchical models of migration destination choice provides a closer fit to observed patterns of migration behaviour than can be achieved using traditional flat processing models.

Similarly, in marketing science, models of brand choice which simulate a hierarchical choice process have been shown to better replicate the actual choices of consumers (Bucklin & Lattin, 1991; Jedidi, Mela & Gupta, 1999). Whilst this is not directly relevant to a spatial choice situation, it does suggest that the human mind uses efficient information storage mechanisms and choice processes in preference to less optimal flat-processing approaches, even in simpler contexts where the number of alternatives is relatively limited.

Summary

This chapter presents a variety of evidence that supports a theory that migration destination choice is an inherently hierarchical process. Three factors are presented which support this conclusion: real world limitations upon the availability of destination information; cognitive storage mechanisms for spatial information, as evidenced by the speed and accuracy of spatial information recall in experiments; and empirical evidence from the application of hierarchical destination choice models.

The research presented in this thesis provides additional evidence in support of such a theory of hierarchical destination choice by calibrating and comparing a number of models of internal migration within Great Britain whose derivation and operationalization are based on the assumption that migration destination choice is a hierarchical process. The next section describes: the models, the choice set definition techniques and the migration and explanatory data that were employed for this purpose.

Chapter Four

Migration data, migration system and explanatory variables

This chapter describes the data used to calibrate the traditional and hierarchical destination choice models presented in this thesis, including: explanatory variables used to describe the potential destinations; observed migration flows against which the models are calibrated; and, the spatial system within which the analysis is performed.

Migration flow data

The data on observed migration was obtained from the Special Migration Statistics (SMS) dataset which was derived from the 1991 Census of Population by the Office for Population Censuses and Surveys (OPCS), (now the Office for National Statistics, ONS). The SMS provides information about independent migrants, which are defined as those individuals aged 16 or over who reported on their Census form that they lived at a different address 12 months prior to Census night, April 21st 1991 (dependent-age children were not considered to be independent migrants). It should be noted that using this definition, multiple moves of an individual within the pre-census year cannot be distinguished, and the related moves of adults within the same household are essentially double-counted, as the data represents them as two independent migration destination choice decisions.

The SMS dataset was accessed from the MIDAS datasets server at Manchester University (now called the MIMAS server, www.mimas.ac.uk) using the SMSTAB access software (MIDAS 1997; Duke-Williams 1995b). The extent to which the SMS disaggregates migrants by age, gender and other socio-economic characteristics is determined by the spatial scale of reporting, with the intention of maintaining the anonymity of all migrants.

The SMS consists of two sets of migration matrices. Set 1 describes flows within and between the 10,933 Census wards in Great Britain (referred to as 'pseudo postcode sectors' in Scotland) and disaggregates migrants by just gender and five-year age groups. Set 2 describes flows within and between the 459 Local Authority Districts and is independently disaggregated by a number of socio-economic variables including: broad age group, ethnic group, marital status, tenure and economic position.

In order to maximise sample sizes and facilitate statistically significant calibration of the models, it was determined that Local Authority District level analysis was the most appropriate spatial scale to use for this analysis. Migrant flows are disaggregated by broad age group, by gender and by marital status.

The flow data used in this research are from: tables 3 and 4 of the SMS Set 2 which disaggregate migrant flows by gender and age group, and by marital status, respectively. The author has also applied some of the techniques applied here to an analysis of gender and marital status variations in migration destination choice (Atkins and Fotheringham, 1999). In order to minimise small-sample effects the majority of this analysis considers the migration behaviour of all migrants aged 16 years and older. Examination of differences in migration behaviour between subgroups is limited to comparisons between three broad age groups: 16-24 years, 25-54 years and 55+ years; between males and females; and between three marital status groups: single, married and widowed/divorced. These broad age categories are designed to group migrants at similar life stages, who it is expected will exhibit some commonality in their migration behaviour. This assumption is supported by previous research by the author showing broadly distinct patterns of inter-district migration behaviour between different broad age group categories (Atkins, 1996). The lower than official retirement age, 55, reflects the decreasing average observed age of retirement and also ensures that the older grouping contains a comparable number of individual migrants to the other two groupings.

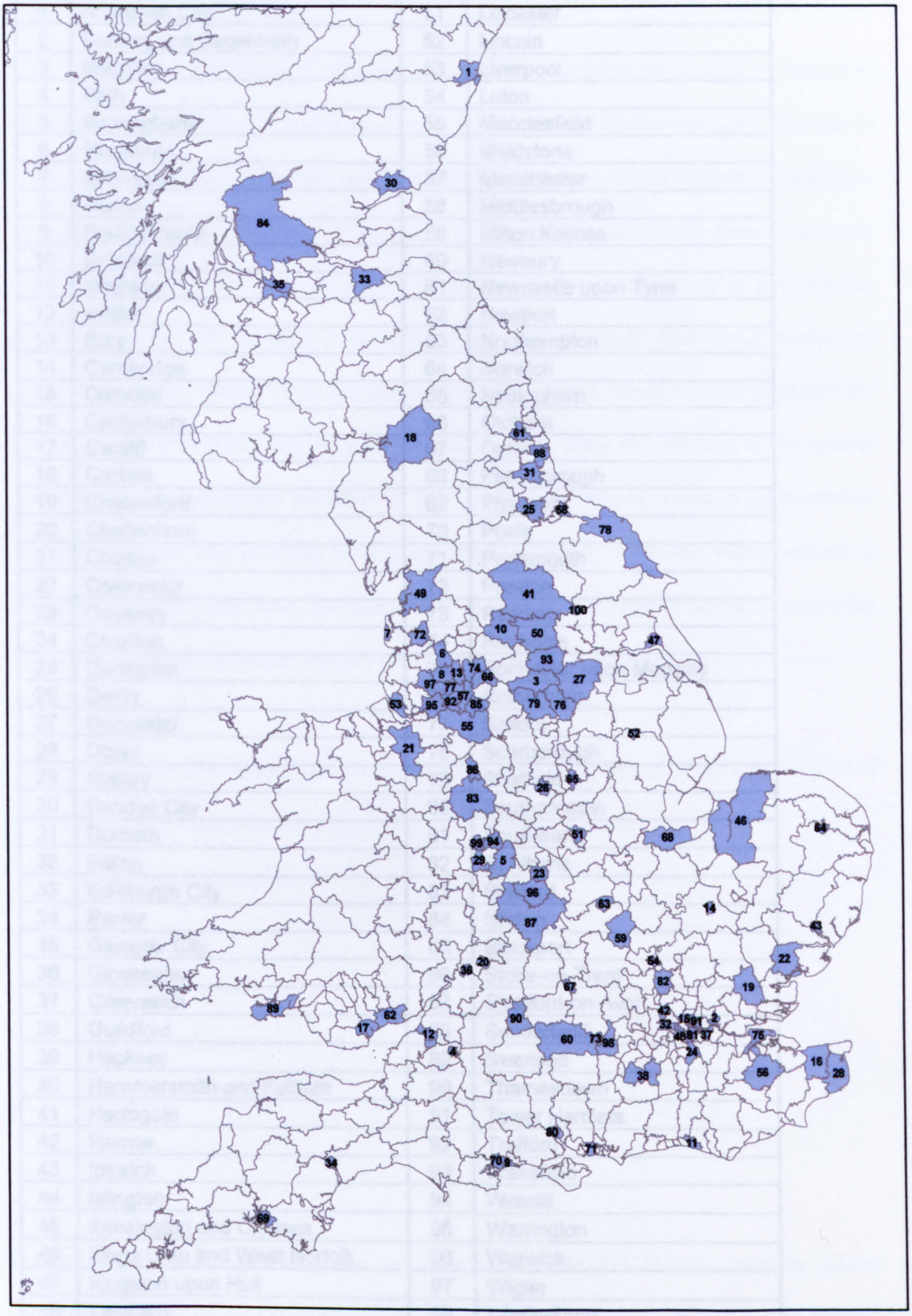
Migration system

As described above, inter-district migration was selected as the most appropriate spatial scale of this analysis. It was further determined that in order to minimise small sample issues, and to ensure that all migration destinations are as 'functionally equivalent' as possible, migration is considered within a system comprising a subset of all Local Authority Districts. The notion of functional equivalence in this context means that each district in the migration system should represent the same 'type' of destination in terms of the migration destination choice process. Specifically, each selected district should have one main settlement that 'defines' that district, rather than being an aggregation of similarly sized but potentially different settlements. Thus, destination differentiation at the district level of the hierarchy is essentially equivalent to comparing major settlements.

One hundred districts were selected based on the following criteria:

- each district should preferably contain just one main settlement, such that this district will be perceived as an individual 'place' rather than as a collection of places
- also districts had to exhibit reasonably high gross out-migration, in order to minimise the small sample problems associated with very small migration flows
- the selected districts should give a fairly representative spatial coverage of the country

This approach was considered preferable to a comprehensive analysis of migration between all districts because some districts exhibit very low out-migration, which would impact statistical significance of any parameter estimates. Also, those districts that contain more than one major settlement were considered inappropriate for this analysis because of the focus in this work on how migrants cluster together and differentiate between major settlements. Such differentiation becomes more complicated and less reliable if some descriptive destination data represents individual settlements whereas other destination data describes the aggregate of a number of settlements. The 100 selected districts are shown in map 4.1 below.



Map 4.1: Map of the migration system of 100 selected districts.

A key providing district names for the districts labelled on map 4.1 is provided in table 4.1.

1	Aberdeen City	51	Leicester
2	Barking and Dagenham	52	Lincoln
3	Barnsley	53	Liverpool
4	Bath	54	Luton
5	Birmingham	55	Macclesfield
6	Blackburn	56	Maidstone
7	Blackpool	57	Manchester
8	Bolton	58	Middlesbrough
9	Bournemouth	59	Milton Keynes
10	Bradford	60	Newbury
11	Brighton	61	Newcastle upon Tyne
12	Bristol	62	Newport
13	Bury	63	Northampton
14	Cambridge	64	Norwich
15	Camden	65	Nottingham
16	Canterbury	66	Oldham
17	Cardiff	67	Oxford
18	Carlisle	68	Peterborough
19	Chelmsford	69	Plymouth
20	Cheltenham	70	Poole
21	Chester	71	Portsmouth
22	Colchester	72	Preston
23	Coventry	73	Reading
24	Croydon	74	Rochdale
25	Darlington	75	Rochester upon Medway
26	Derby	76	Rotherham
27	Doncaster	77	Salford
28	Dover	78	Scarborough
29	Dudley	79	Sheffield
30	Dundee City	80	Southampton
31	Durham	81	Southwark
32	Ealing	82	St Albans
33	Edinburgh City	83	Stafford
34	Exeter	84	Stirling
35	Glasgow City	85	Stockport
36	Gloucester	86	Stoke-on-Trent
37	Greenwich	87	Stratford-on-Avon
38	Guildford	88	Sunderland
39	Hackney	89	Swansea
40	Hammersmith and Fulham	90	Thamesdown
41	Harrogate	91	Tower Hamlets
42	Harrow	92	Trafford
43	Ipswich	93	Wakefield
44	Islington	94	Walsall
45	Kensington and Chelsea	95	Warrington
46	Kings Lynn and West Norfolk	96	Warwick
47	Kingston upon Hull	97	Wigan
48	Lambeth	98	Wokingham
49	Lancaster	99	Wolverhampton
50	Leeds	100	York

Table 4.1: The 100 districts selected for the migration system for this analysis.

Destination characteristics

The selection of appropriate destination characteristic variables for models of migration destination choice may well be the most contentious topic of migration research. The task of selecting appropriate explanatory variables has persisted as various modelling paradigms have come in and out of vogue. With respect to this variable selection problem, the reader should recall that whilst it is desirable to model migration as accurately as possible, the central focus of this research is to demonstrate that inherently hierarchical models of migration provide a better approximation to observed migrant behaviour than do traditional non-hierarchical models given the same explanatory data. Thus the selection of the optimal combination of explanatory variables is not key to this research, rather a representative sample of explanatory variables have been used all of which have been widely applied to migration analysis and have generally been shown to significantly relate to migration behaviour, (Atkins and Fotheringham, 1999; Boyle, 1993; Congdon, 1988).

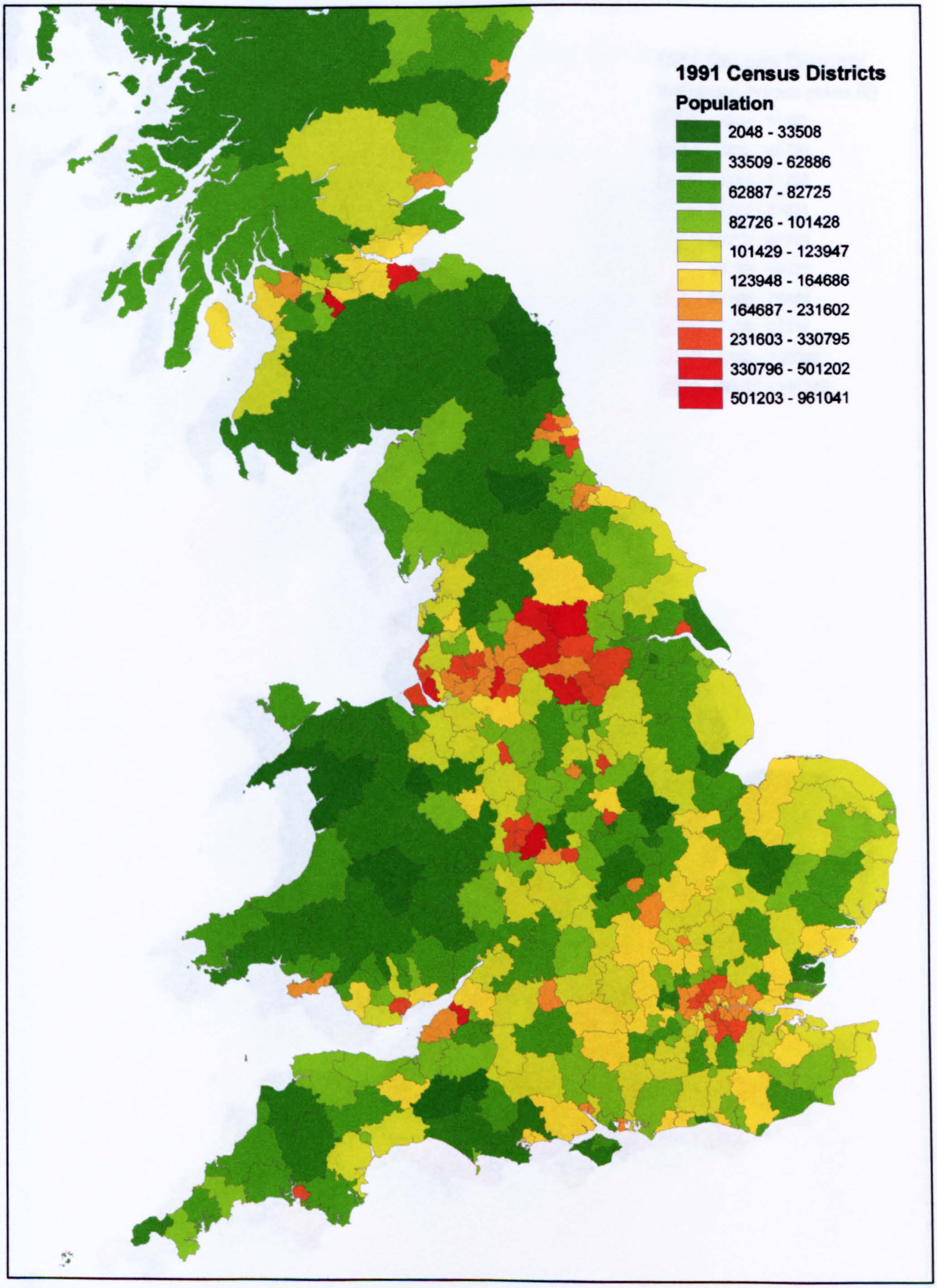
The various migration models calibrated in this research all make use of the following set of variables describing the destination districts in the migration system:

- origin-destination separation
- resident population
- average house price
- social class structure
- housing tenure structure
- unemployment rate

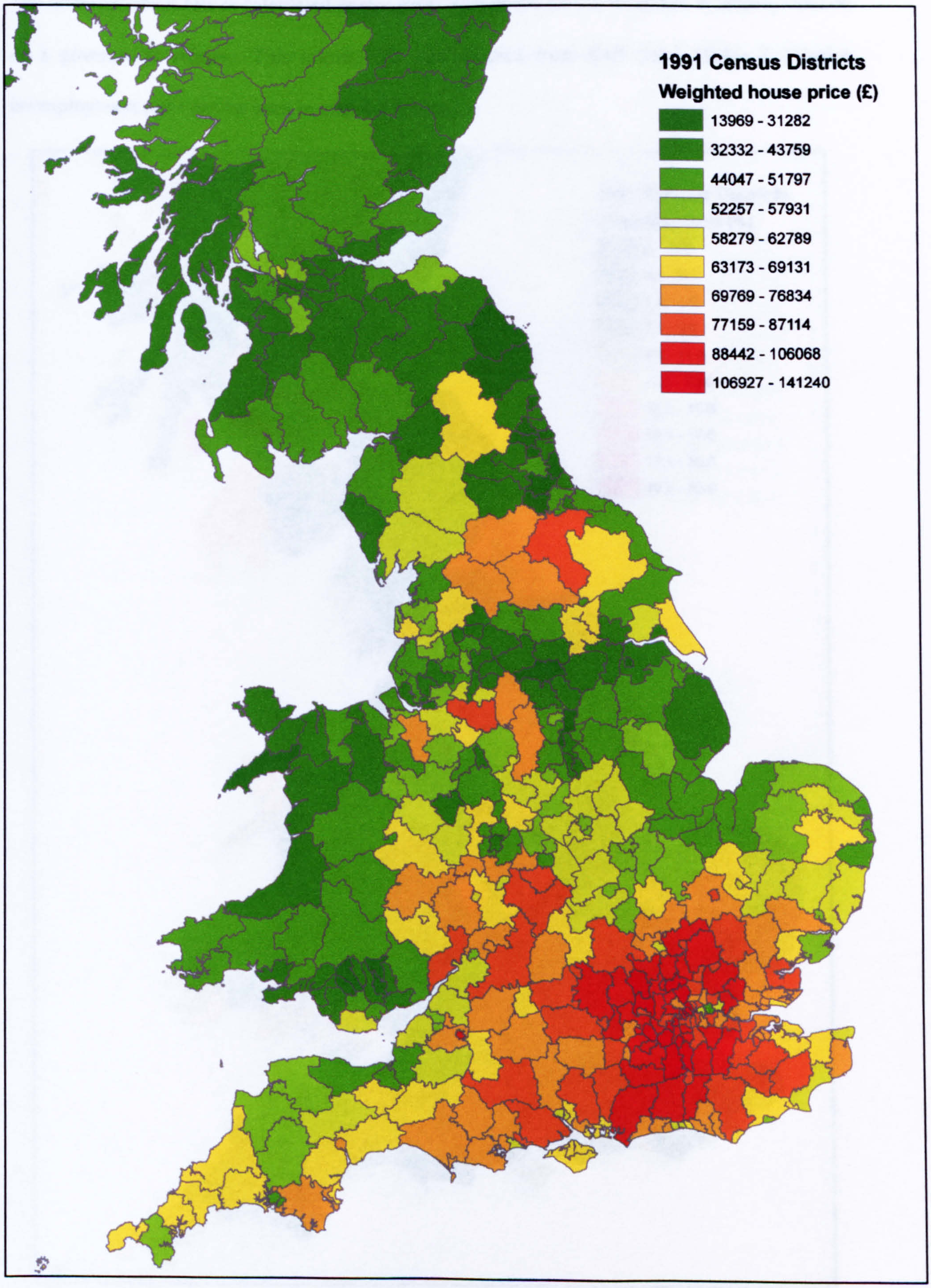
Origin-destination separations were calculated as straight-line distance between district centroids. These district centroids were derived from corrected and population-weighted ward centroids (Atkins *et al*, 1993). The use of straight-line origin-destination separations in migration models has been criticised as potentially not accurately representing migrants' perception of distance, and numerous alternatives have been proposed such as rail/road distance or travel time (Atkins, 1996). However, given the difficulty in calculating better separation measures, straight-line distance was considered to be an adequate surrogate for perceived separation. Furthermore, because the incorporation of hierarchical choice into migration models is intrinsically linked to the spatial structure of the migration system, the introduction of additional non-intuitive spatial structure through a complex separation variable would complicate the interpretation of a model's parameter estimates. If the use of a straight-line separation variable has any significant shortcomings these will likely become apparent when the residual flows from model calibrations are considered.

The usually resident population of each district was obtained from table 1 of set 1 of the 1991 Census Small Area Statistics (SAS) dataset. This includes those people normally resident in each district but absent on Census night but excludes visitors. The populations of the selected districts can be compared on map 4.2 below.

House price data were obtained from a high street building society based upon property sales in 1990, and are weighted by property type to account for variation in housing stock between districts. The weighted property prices for all districts can be compared in map 4.3 below.

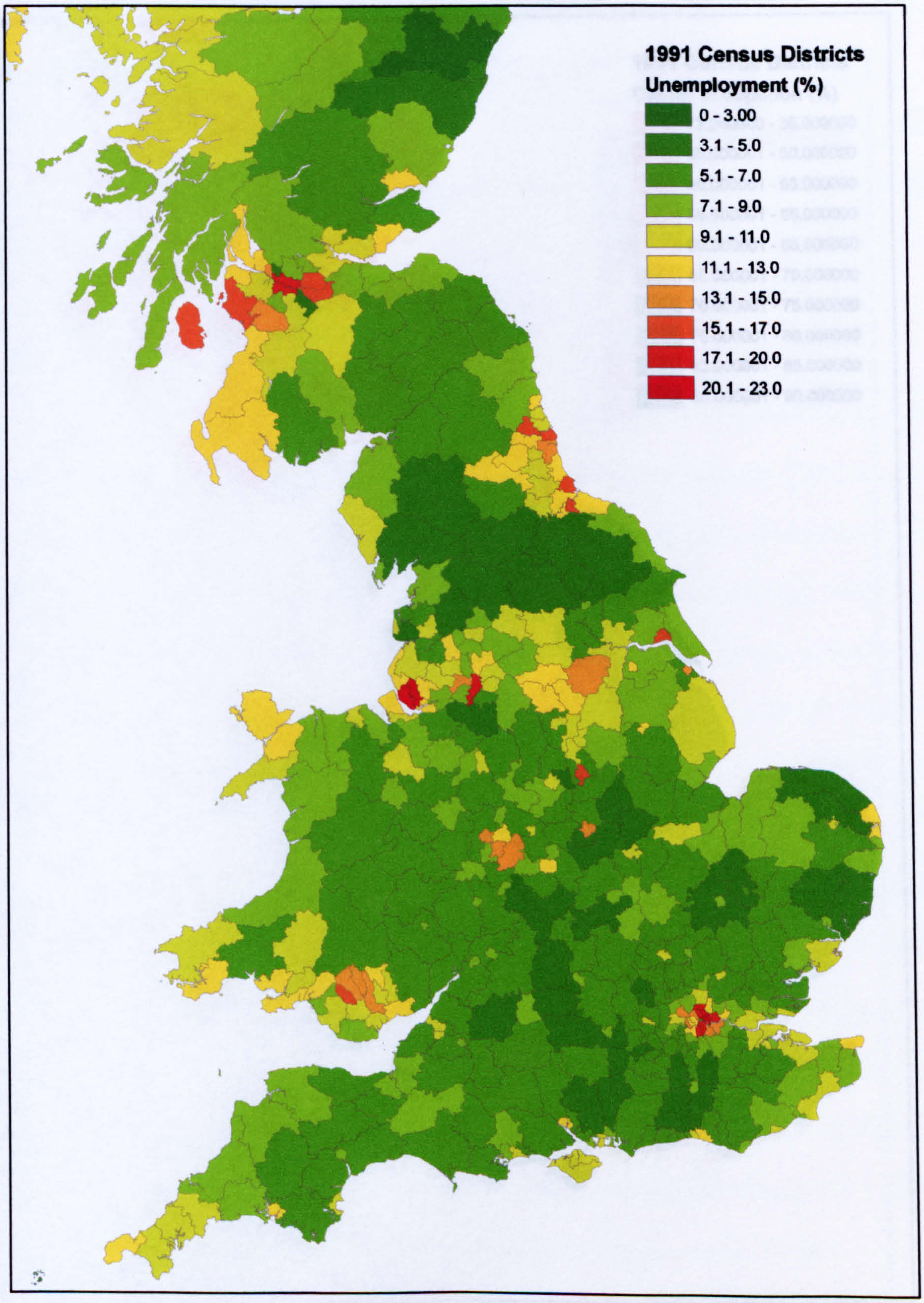


Map 4.2: Usually resident populations of local authority districts.



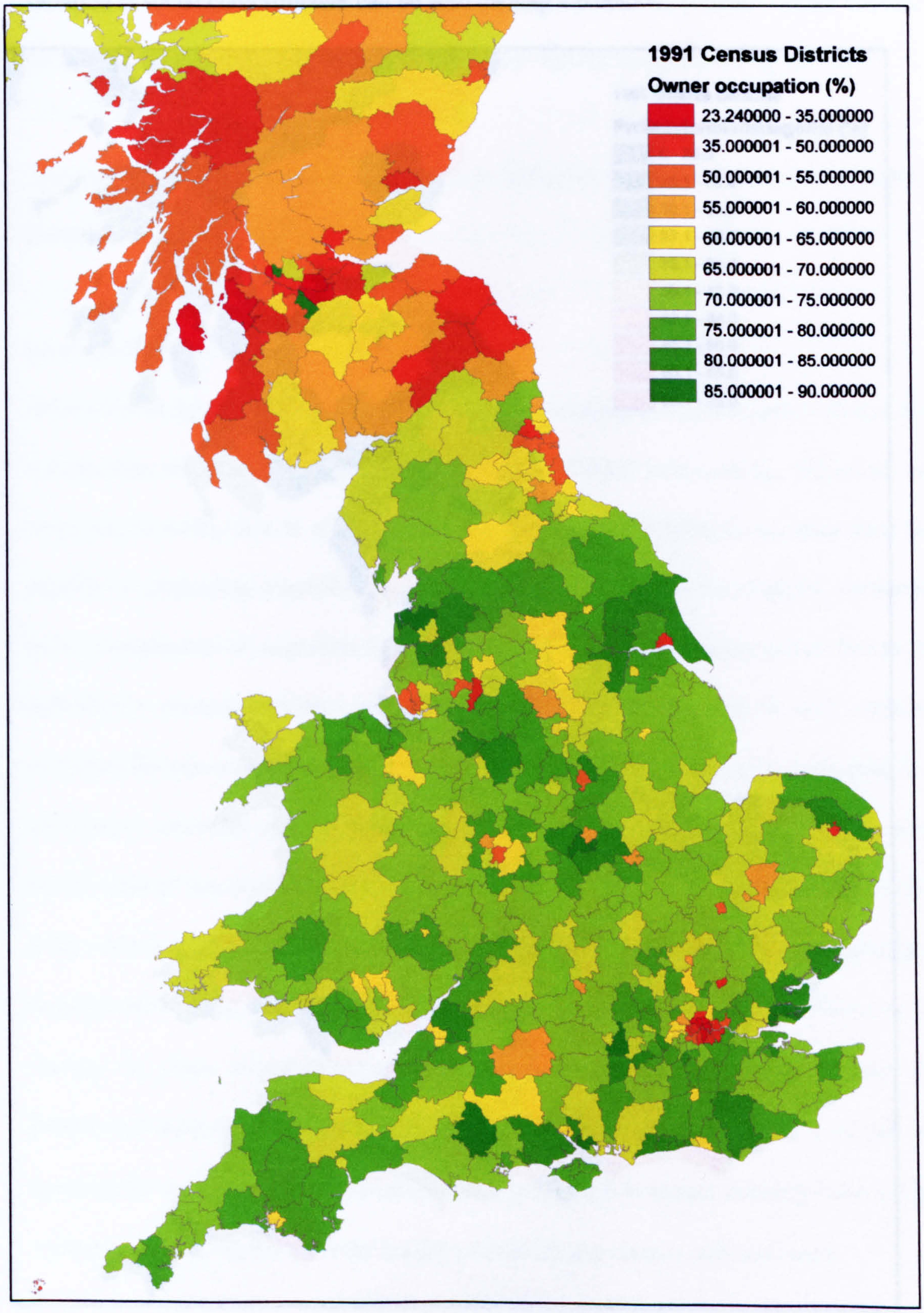
Map 4.3: Weighted house prices for local authority districts.

The unemployment rate is calculated as that proportion of the labour force not in employment or on a government scheme. This information was obtained from SAS Set 1, Table 8. District unemployment rates can be seen in map 4.4 below.



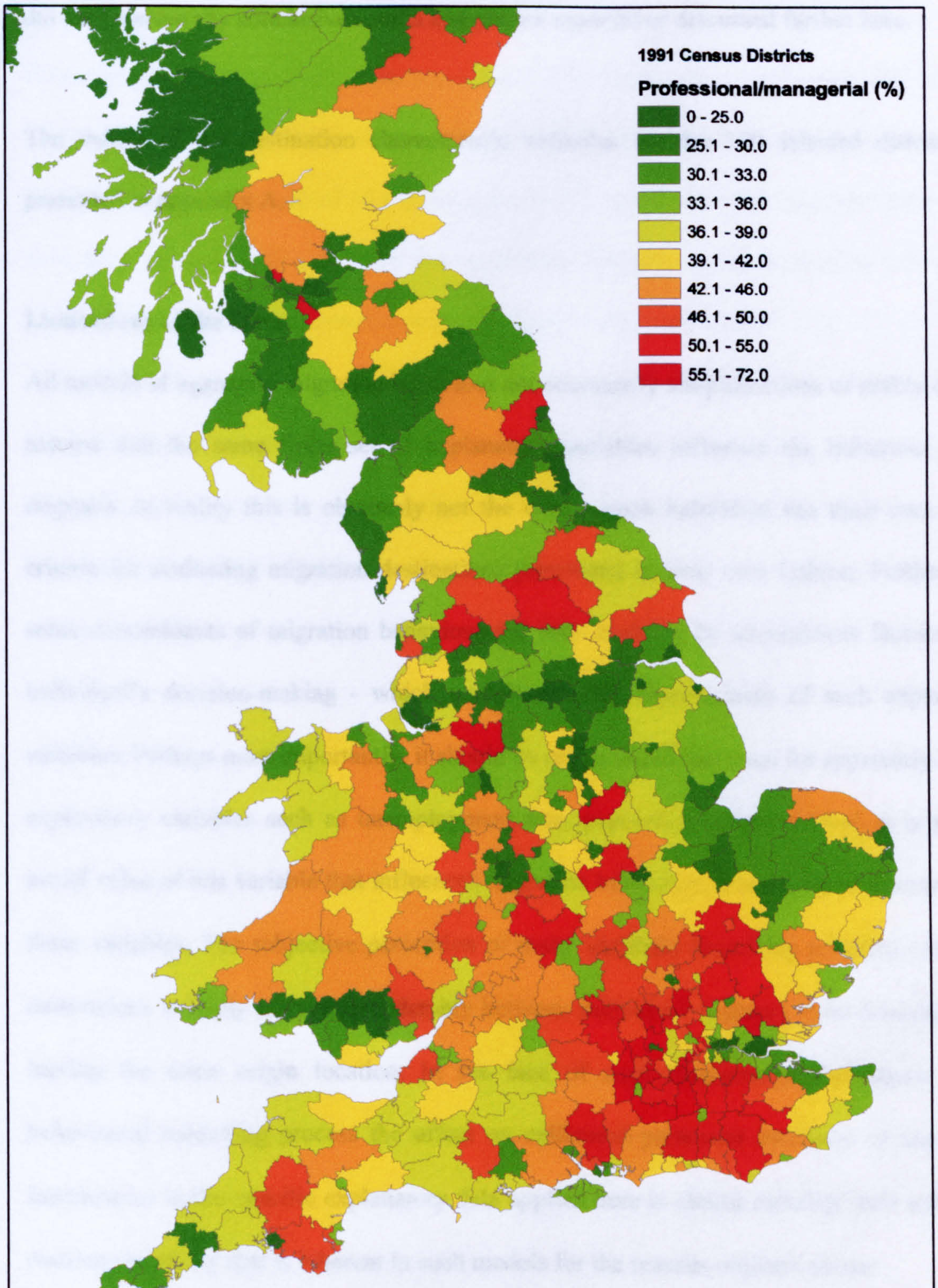
Map 4.4: Unemployment rates for local authority districts.

The tenure structure of each district is represented by the percentage of households in that district that are owner-occupied. This data was obtained from SAS Set 1, Table 20. Owner occupancy rates for all districts can be seen in map 4.5 below.



Map 4.5: Owner occupancy rates for local authority districts.

The social class structure of each district is represented by the percentage of heads of households who are in professional and senior managerial employment, which corresponds to social classes I and II in Census terminology. This data was obtained from SAS Set 1. Spatial variation in social class structure can be seen on map 4.6 below.



Map 4.6: Social class variable for local authority districts.

Some exploratory model calibrations were made during the course of this research using dummy variables to represent destination contiguity and whether destinations are London Boroughs. These two dummy variables were generated using GIS analysis of digitised district boundaries and manual inspection of OPCS district codes, respectively. These calibrations did not produce any conclusive results and are not reported or discussed further here.

The values of the destination characteristic variables for the 100 selected districts are presented in appendix A.

Limitations of the Data

All models of aggregate migration behaviour are necessarily simplifications of reality as they assume that the same finite set of explanatory variables influence the behaviour of all migrants. In reality this is obviously not the case – each individual has their own set of criteria for evaluating migration destinations prioritised in their own fashion. Furthermore, some determinants of migration behaviour are very likely to be unconscious factors in an individual's decision-making - which complicates the identification of such explanatory variables. Perhaps most importantly, it should be remembered that even for apparently simple explanatory variables such as unemployment rate, population or crime level, it is not the actual value of this variable that influences migration behaviour, it is migrants' *perception* of those variables. The subjective perception of many variables describing potential migration destinations is likely to vary considerably between individual migrants, even between those leaving the same origin location. In the face of such fundamental limitations of the behavioural modelling process the effect on calibrated parameter estimates of any minor inaccuracies in the specific explanatory data applied here is almost certainly well within the random variability that is inherent in such models for the reasons outlined above.

Given these constraints the best that can be attempted is to include a selection of objectively accurate explanatory variables that likely influence the decision-making of the majority of migrants (preferably influencing them in a consistent manner). That is what is undertaken and attempted in this thesis.

Fortunately, we are primarily concerned here with the relative, rather than the absolute performance of a set of hierarchical and non-hierarchical models of migration destination choice. Thus, the impact of the limitations in the empirical data and the behavioural modelling process outlined above are minimized because these limitations affect both hierarchical and traditional non-hierarchical models to the same extent.

The remaining two chapters in this section describe the derivation of the migration destination choice models used in this research, and the methods of generation of the hierarchical destination choice sets required by these models.

Chapter Five

Migration destination choice models

Introduction

The arguments supporting a theory of hierarchical migration destination choice were presented in chapter 3. One category of evidence for such a theory is the successful application of migration models whose derivation is based upon assumptions of hierarchical decision-making. This chapter introduces four such models – the competing destinations model, the discrete nested logit model, the weighted nested logit model and the hybrid weighted nested logit model. The results obtained from the calibration of these models are presented in chapters 7 through 10.

Aspatial information processing

It is useful here to recall the formulation of the traditional migration destination choice model, as discussed in chapter 2. This model is the share logit model of discrete choice that predicts the number of migrants to each potential destination by multiplying the predicted likelihood of each potential destination being selected by the total number of migrants. This model is based on the assumption that migrants are rational decision makers who will select the destination that offers them the highest utility.

The formulation of the model, repeated from chapter 2, is shown in equation 5.1 below. Note that the utility is expanded into its constituent explanatory variables in this formulation:

$$M_{ij} = M \frac{exp\,(O_{i1}^{\alpha 1} + O_{i2}^{\alpha 2} + ... + O_{if}^{\alpha f} + D_{j1}^{\lambda 1} + D_{j2}^{\lambda 2} + ... + D_{jg}^{\lambda g})}{\sum_k exp\,(O_{i1}^{\alpha 1} + O_{i2}^{\alpha 2} + ... + O_{if}^{\alpha f} + D_{k1}^{\lambda 1} + D_{k2}^{\lambda 2} + ... + D_{kg}^{\lambda g})} \qquad (Eq.5.1)$$

Where:

M_{ij} is the migration from origin i to destination j
M is the total number of migrants
$O_{i1}^{\alpha 1}$ is parameterised explanatory variable 1 describing origin i
$D_{j1}^{\lambda 1}$ is parameterised explanatory variable 1 describing destination j

Equation 5.1: Full formulation of unconstrained global traditional logit model.

Equation 5.1 represents the most general version of the traditional logit model of migration destination choice. It is a global model that estimates migration from all origins to all destinations. Whilst this type of global model is calibrated in this research, the primary focus is on more local analysis based on the results from origin-specific model calibrations. The simplified formulation of the origin-specific production-constrained logit model is presented in equation 5.2.

$$M_{ij} = M_i \frac{exp\,(D_{j1}^{\lambda 1}\, D_{j2}^{\lambda 2} ... D_{jg}^{\lambda g})}{\sum_k exp\,(D_{k1}^{\lambda 1}\, D_{k2}^{\lambda 2} ... D_{kg}^{\lambda g})} \qquad (Eq.\ 5.2)$$

Where:

M_{ij} is the migration from origin i to destination j
M_i is the total out-migration from origin i
$Dj_1^{\lambda 1}$ is parameterised explanatory variable 1 describing destination j

Equation 5.2: Production-constrained origin-specific traditional logit model.

Two statistical properties of the traditional model were mentioned in chapter 2 which make that model inappropriate for predicting spatial decisions such as migration destination choice: Independence from Irrelevant Alternatives (IIA), and, Regularity.

The IIA property essentially means that the ratio of the likelihoods of selection of any two potential destinations cannot be affected by the addition of an additional destination. This is intuitively incorrect in a spatial choice context where the location of any additional alternative relative to existing alternatives will obviously have a different effect on the likelihood of selection of those areas. The exact nature of this effect is not so intuitive – introduction of a very appealing destination could increase migration to its neighbours by making the general area more attractive to migrants (an agglomeration effect). Alternatively, the direct competition that the new alternative poses to its neighbours could reduce their likelihood of selection. As will be discussed below, there is theoretical as well as empirical evidence that competition effects become more significant than agglomeration effects as the number of alternatives in a cluster increases.

Similarly, the property of regularity – the fact that the likelihood of selection of any particular destination cannot be increased by the introduction of another alternative – is not appropriate in a context where location is a key characteristic of the various alternatives. For instance, if following the introduction of a new alternative, the agglomeration effects discussed above proved to be more powerful that the competition effects, then the existence of that new alternative would indeed increase the popularity of its neighbours.

Another interesting consideration is that the spatial location of a new alternative may not only make other destinations appear more or less attractive to decision makers, it can also affect the likelihood that any particular destination will even be included in a particular migrant's choice set. For instance, it is quite possible, indeed likely, that not all of the individual members of a closely clustered group of potential destinations at one end of the country would be known about and individually evaluated by migrants leaving origins at the other end of the country. Migrants may well not have enough detailed 'local' knowledge about distant districts to be able to differentiate all the members of the destination cluster, in which case only a subset of the destinations in that cluster are likely to exist in any individual migrant's choice set.

This consideration of whether a particular destination exists in a particular migrant's choice set is particularly interesting because, whilst derived from a behavioural perspective, i.e. consideration of the actual decision-making process at an individual level, it also builds unconscious or subconscious behaviour into the modelling process. The main determinants of whether a destination will receive attention from a migrant are the amount of destination information a migrant has available and the extent to which the migrant increasingly underestimates the size of destination clusters as they become larger. Migrants are unlikely to search out more information about places that they do not know exist, and the vast majority of migrants would be likely to claim that they do not underestimate the size of larger destination clusters. So despite the existence of 'behavioural factors' in the choice process it is interesting to note that these characteristics would be very hard to analyse using more traditional behavioural research methods such as migrant interviews.

There are also a number of practical assumptions inherent in the traditional flat-processing logit model that are intuitively unreasonable. In particular, the very high number of alternatives in the decision-making process and the lack of accurate and sufficient information describing each of those alternatives. Whilst the distance variable in a traditional migration model can be considered a likely determinant of the quantity and quality of information that a migrant will have about a particular destination, the same accurate and objective explanatory variables are included in the model for each and every destination regardless of their distance from the origin location. Intuition suggests that it would be a more accurate reflection of reality if more precise and detailed destination data were included in a model for more proximal destinations that for more distant locations. Or perhaps more appropriately, more accurate and detailed information was included in a model for the destinations within the region(s) that a migrant had expressed a preference for. That, of course, would be a hierarchical and not a flat-processing model of migration, and is somewhat akin to the operation of the nested logit model that will be discussed further below.

It has also been presented by Fotheringham (1981) that the spatial structure in a migration destination choice set causes bias in the parameter estimates for the distance variable in models that take no account of this spatial structure. Fotheringham presents several existing theories that could explain this bias, including growth centre theory. Fotheringham states that growth centre theory suggests that the growth of smaller destinations will be related to its interactions with larger destinations, which that amount of interaction will be related to those smaller areas' locations relative to those larger areas – i.e. the spatial structure of the destination choice set – causing misspecification of the model if this spatial structure is not included in the model. (Fotheringham, 1982)

The various shortcomings of the traditional flat-processing approach to migration destination choice modelling presented above demonstrate the need for the relative location of potential migration destinations to be represented in models of migration destination choice. The following sections introduce a number of models that use different approaches to incorporating this relative destination location information into the migration modelling process.

Spatial Information Processing

The central proposition of this thesis is that due to the limited destination information available to migrants and their finite information processing capacity, a hierarchical, step-wise decision-making process is more appropriate than a 'flat processing' approach when selecting a migration destination. This suggests that the selection of a house to move to is the result of previous preference decisions at larger spatial scales, such as regions, counties, cities, neighbourhoods...etc. All such decision-making is based upon conscious or unconscious clustering of destinations into regions, or city-areas. Four models are applied in this research employing various approaches to the incorporation of such destination clustering into the modelling process: the competing destinations model, the discrete nested logit model, the weighted nested logit model and the hybrid weighted nested logit model. Each is briefly described here:

The competing destinations model uses an accessibility variable to represent the likelihood (based on spatial location relative to other destinations) of a particular destination being cognized by migrants as being in a larger cluster of destinations.

The discrete nested logit model imposes spatial structure on the destination choice set by categorizing destinations into a discrete regionalization, and employs a regional utility variable to determine the extent to which migrants' destination choices make use of regional- or cluster-level characteristics. This regionalization is generated using a quasi-random technique based on information maximizing principles that take into account the relative location of all the destinations.

The weighted nested logit model follows a similar approach to the discrete nested logit model by applying a pre-determined regionalization on the destination choice set. However, in this case the regionalization is defined by a matrix that represents the likelihoods that any two destinations will be perceived by a migrant to be within the same region. Though one could debate whether an individual migrant would have a discrete or probabilistic internal mental map of the destination clusters, it is intuitively more attractive to use the probabilistic approach for aggregate migration destination choice modelling, because this is likely to provide a better representation of the wide range of different mental maps constructed by the many individual migrants leaving each origin.

The hybrid weighted nested logit model makes use of the accessibility variable from the competing destinations model and also the weighted regional utility variable from the weighted nested logit model. Comparison of results from the competing destinations and weighted nested logit models suggests that whilst both models provide better predictive ability that the traditional model, their patterns of improvement appear to differ sufficiently to justify the derivation and application of a hybrid model combining the hierarchical approaches of both models.

The process by which the clustering likelihood variable, and the discrete and probabilistic regionalizations are generated is described in detail in the next chapter.

Competing destinations model

The competing destinations model is based upon the assumption that migrants select migration destinations hierarchically, meaning that they will mentally group the potential destinations. Research in the field of perceptual psychology has demonstrated that perception of group size is non-linear, such that the size of larger groups or clusters is underestimated, (Stevens, 1975). From this finding, one would anticipate that a region comprising 20 potential migration destinations will not be perceived to provide twice the utility of a region of 10 'identical' destinations, and thus in a rational utility-maximizing choice process the larger region will be selected over the smaller region less often than its true scale would merit. It follows that an individual destination within a larger group of destinations is less likely *ceteris paribus* to be selected than an individual destination in a smaller group of destinations. The competing destinations model compensates for this psychophysical trait by including an accessibility variable that represents the likelihood of any particular destination being cognized within a larger group of destinations. In order to achieve this, the accessibility variable is calculated based on the relative separation and sizes of all potential destinations.

The accessibility of each destination is calculated according to the formula presented in equation 5.3 below:

$$A_i = \sum_{\substack{j=1 \\ j \neq i}} (P_j^{\alpha} / d_{ij}^{\beta}) \qquad \textit{(Eq. 5.3)}$$

Where:

A_i is the accessibility of district i
P_j is the population of destination j
d_{ij} is the distance from origin i to destination j

Equation 5.3: General formulation of the accessibility variable.

The α and β exponents on the population and population and distance variables in equation 5.3 determine how much the accessibility value of an area is influenced by the size and proximity of its neighbours.

Appropriate values for these exponents were determined through examination of the goodness-of-fit and parameter estimate significant from extensive exploratory calibrations of competing destinations models with accessibility variables derived from a range of distance and population exponents. This process, which is documented in more detail in chapter 6, led to the values of 2.5 and 1.5 being selected for the population and distance exponents, respectively, for the purposes of the analysis reported here.

Because population and distance appear in the numerator and denominator, respectively, in equation 5.3, and have positive exponents, higher accessibility values indicate that an area is proximal to more and larger destinations. Thus, the higher the accessibility of a destination, the more likely it is to be cognized within a larger group of larger destinations. This being the case, the psychophysical tendency, noted above, to increasingly underestimate the membership of larger groups, will manifest itself as negative estimates for the accessibility variable parameter when the competing destinations model is calibrated.

The competing destinations model can be formulated as shown in equation 5.3 below.

$$M_{ij} = M_i \frac{exp\ (D_{j1}^{\lambda 1}\ D_{j2}^{\lambda 2}\ ...D_{jg}^{\lambda g}\ A_j^{a})}{\sum_k exp\ (D_{k1}^{\lambda 1}\ D_{k2}^{\lambda 2}\ ...D_{kg}^{\lambda g}\ A_k^{a})} \qquad (Eq.\ 5.4)$$

Where:

M_{ij} is the migration from origin i to destination j
M_i is the total out-migration from origin i
$D_{j1}^{\lambda 1}$ is parameterised explanatory variable 1 describing destination j
A_j^{a} is parameterised accessibility variable for destination j

Equation 5.4: The competing destinations model.

It can be seen from equation 5.4 that, despite the difference in their derivation, the competing destinations model is very similar structurally to the traditional logit model (see equation 5.1). Upon reading the competing destinations model formulation presented in equation 5.4 it is tempting to consider the accessibility variable as just another characteristic variable contributing to a destination's utility. However, given the theoretical basis for this variable, it is more appropriate to think of the accessibility variable as acting to define the choice set, or more specifically, the likelihood of any particular destination appearing in any particular migrant's choice set. It is potentially useful to consider the more general formulation of the competing destinations model to be as shown in equation 5.5 below (Fotheringham et al, 2000).

$$M_{ij} = M_i \frac{exp\ (V_{ij})\ P(j \in C_i)}{\sum_k exp\ (V_{ik})\ P(k \in C_i)} \qquad (Eq.\ 5.5)$$

Where:

M_{ij} is the migration from origin i to destination j
M_i is the total out-migration from origin i
V_{ij} is a vector explanatory variables describing destinations
$P(j \in C_i)$ is probability that destination j is in the choice set of a migrant from origin i

Equation 5.5: Competing destinations model as a traditional logit with a choice set modifier.

Within the context of this formulation the term A_j^a from equation 5.3 can be considered to represent the $P(j \in C_i)$. Stevens' psychophysical observation that individuals increasingly underestimate the size of larger clusters suggests that *ceteris paribus* any individual destination in a larger destination cluster is more likely to be overlooked by any particular migrant than any particular destination in a smaller cluster. Or in other words, the probability of a destination in a larger destination cluster appearing in any particular migrant's conscious choice set is smaller than the probability of a destination in a smaller destination cluster being in that same choice set. Given that the accessibility variable represents the likelihood of a destination being cognized within a larger group of larger destinations, the parameterized accessibility term A_j^a in equation 5.4 can be interpreted as representing the likelihood $P(k \in C_i)$ of any particular destination appearing in a particular migrant's choice set.

There are numerous applications of the competing destinations model to a variety of spatial interaction problems. Guldmann (1999, 2000) used the model to predict international telecommunications flows, and concluded that inclusion of variables representing spatial structure effects in the choice set do improve model predictions. Mitchelson and Wheeler (1994) applied the model to the prediction of domestic and international information flows, as represented by data on FedEx shipments.

Thus the competing destinations model can be considered to be a traditional logit with a destination choice set that is modified through the inclusion of an accessibility variable that takes account of the spatial structure of the various destinations. The nested logit family of migration destination choice models described below employ a different mechanism of choice set modification.

Discrete nested logit model

The IIA (Independence from Irrelevant Alternatives) property of the traditional migration destination choice model, mentioned above, is intuitively inappropriate to the modelling of spatial decisions. However, even in aspatial decision-making situations IIA is not always an acceptable assumption. For instance, when modelling brand choice of, say, chocolate bars, the introduction of an additional brand that is almost identical to an existing brand will likely affect the popularity of the very similar alternative more than dissimilar brands. In such aspatial situations where there is an inherent grouping or hierarchy in the choice set, the nested logit model is often applied. The nested logit model uses an a-priori classification of the choice set into a multi-level hierarchy and introduces a group utility variable(s) into the model that allows consideration of the relative attractiveness of the groups of alternatives as well as the relative utility of individual alternatives within those groups.

The same approach is applied in this analysis which groups the migration destination choice set based on the relative spatial situation of the various alternatives. A comprehensive two-level hierarchy (simply a grouping of all districts into regions) is generated to represent migrants' likely mental grouping of the potential migration destinations. This regionalization forms the basis for the calculation of a regional utility variable which is used in origin-specific calibrations of the discrete nested logit model.

The derivation of these discrete regionalizations is described in detail in chapter 6 along with maps of some example regionalizations. In summary, however, a semi-random, non-deterministic algorithm is used to produce an allocation of each and every district to exactly one of an appropriate number of regions such that regional information variance is minimized relative to a particular migration origin. The nature of regionalization process typically results in discrete regionalizations that contain smaller regions closer to the migration origin under consideration, and larger regions further away from that origin. This is evident from example maps in the next chapter (maps 6.3 and 6.4) which show smaller regions in northern England, close to Leeds, and larger regions in the far south and north of the country.

The discrete nested logit model can be formulated as shown in equation 5.6 below.

$$M_{ij} = M_i \frac{exp\,(D_{j1}^{\lambda 1}\, D_{j2}^{\lambda 2} \ldots D_{jg}^{\lambda g}\, DR_J^{a})}{\sum_k exp\,(D_{k1}^{\lambda 1}\, D_{k2}^{\lambda 2} \ldots D_{kg}^{\lambda g}\, DR_K^{a})} \qquad (Eq.\ 5.6)$$

Where:

M_{ij} is the migration from origin i to destination j
M_i is the total out-migration from origin i
$D_{j1}^{\lambda 1}$ is parameterised explanatory variable 1 describing destination j
DR_J^{a} is parameterised discrete regional utility for region J containing destination j

Equation 5.6: The discrete nested logit model.

It can be seen from equation 5.6 that this model also has a similar structure to the traditional logit model, but with the addition of a regional utility variable. Like the accessibility variable in the competing destinations model, the inclusion of this regional utility variable in the discrete nested logit model essentially acts as a way of modelling the likelihood of any particular destination appearing in a particular migrant's destination choice set. This makes particular sense if one considers destination choice as a step-wise process whereby a migrant selects a region and then a sub-region/city, neighbourhood...etc. If two otherwise identical destinations have very different regional utility values, then it is much more likely that the destination in the 'more desirable region' will be evaluated and potentially selected by a migrant than the destination in the less attractive region.

A region's utility variable is the sum of the values of the utility variable for all destinations allocated to that region. Thus all potential destinations within the same region will share the same regional utility value. The discrete regional utility is formulated in equation 5.7 below.

$$DR_K = \sum_{k \in K} (D_{k1}^{\lambda 1} D_{k2}^{\lambda 2} \ldots D_{kg}^{\lambda g}) \qquad \textit{(Eq. 5.7)}$$

Where:

DR_K is the discrete regional utility of region K

$D_{k1}^{\lambda 1}$ is parameterised explanatory variable 1 describing destination k

Equation 5.7: Discrete regional utility.

In operational terms, the calibration of the discrete nested logit model can be performed in two ways:

- direct maximization of the nested logit model's likelihood function,
- sequential calibration of two logit models: first, a simple traditional logit model is calibrated and the resulting parameter estimates are used to calculate values of the regional utility variable which is then used as an addition variable in a second logit model calibration.

The latter approach is much more computationally tractable and has been shown to provide a very good approximation to the maximization of the likelihood function for hierarchical choice sets with three or fewer levels where there are at least five alternatives in each group at each level of the hierarchy (Ben-Akiva and Lerman, 1987). These requirements for accurate sequential calibration are enforced during the discrete regionalization generation process by imposing checks to ensure each generated regionalization comprises sufficient regions and also that there are sufficient migration destinations within each region.

There have been some applications of the discrete nested logit model to migration research, but these typically make use of the additional dimension(s) in the choice hierarchy to differentiate between types of migrants or types of destination rather than as an additional spatial characteristic (Newbold, 1996; Cameron, 2000). A notable exception is Lin and Xie (1998) who employ a discrete nested logit approach to model interstate migration within the

United States. In their analysis the higher level of the spatial hierarchy are broad geographical regions, and they identify a regional holding power variable, which is equivalent to the regional utility variable in the discussion.

Another approach to hierarchical modelling termed hierarchical linear modelling (HLM) has some similarities with the discrete nested logit model, insofar as when they are applied in a spatial context both models require a predefined discrete spatial hierarchy. HLMs were originally developed by researchers investigating educational effectiveness of schools, and HLMs have proved effective at separating the variance in a dependent variable (such as student performance) that results from explanatory variables at different levels in the model's hierarchy (i.e. school and Local Education Authority might be two such levels in a hierarchy) (Goldstein, 1987). Whilst some efforts have been made to apply these models in a spatial context, the complexities of their interpretations, which is essentially multilevel ANOVA, have meant they have not been widely applied by spatial scientists, and the author is aware of no examples of their use in spatial interaction modelling (Jones, 1991).

Weighted nested logit model

Whilst the discrete nested logit model does incorporate the relative locations of migration destinations into the modelling process, it does impose an intuitively unacceptable assumption – that the heterogeneous mental maps of all migrants leaving a particular origin can be effectively represented by a single discrete allocation of destinations to regions.

The weakness of the discrete nested logit model's dependency on a single rigid regionalization becomes evident when one examines the sensitivity of the model's parameter estimates to the specific discrete regionalization against which it is calibrated. Detailed comparisons of results obtained from calibrating discrete nested logit models against different discrete regionalizations are presented in chapter 8.

In an attempt to overcome this problem, probabilistic concepts were applied to derive the weighted nested logit model, which is calibrated against a non-discrete probabilistic regionalization of migration destinations. Specifically, rather than discretely defining a destination IS or IS NOT in the same region as another, a probabilistic regionalization defines the likelihoods that any particular pair of destinations will be cognized as being within the same region by any particular migrant.

The probabilistic regionalization is represented by a 459-square matrix, each element of which corresponds with a specific pairing of districts. The value of each element in the matrix represents the likelihood of a particular pair of districts being considered by a migrant to be within the same region. A detailed description of the generation of probabilistic regionalizations is presented in the chapter 6.

The elements of the matrix representing a probabilistic regionalization will range in value from 0 to 1, with a value close to 1 being indicative of a pair of districts that occur in the same region in most of the many discrete regionalizations that were used to generate the probabilistic regionalization, and which, intuition suggests, will therefore be cognised as being in the same region by the majority of migrants.

$$M_{ij} = M_i \frac{exp\ (D_{j1}^{\lambda 1}\ D_{j2}^{\lambda 2}\ ...D_{jg}^{\lambda g}\ WR_j^{a})}{\sum_k exp\ (D_{k1}^{\lambda 1}\ D_{k2}^{\lambda 2}\ ...D_{kg}^{\lambda g}\ WR_k^{a})} \qquad \textit{(Eq. 5.8)}$$

Where:

M_{ij} is the migration from origin i to destination j
M_i is the total out-migration from origin i
$D_{j1}^{\lambda 1}$ is parameterised explanatory variable 1 describing destination j
WR_j^{a} is parameterised weighted regional utility for destination j

Equation 5.8: The weighted nested logit model.

It is clear from equation 5.8 that the formulation of this model is almost identical to that for the discrete nested logit model (see equation 5.6). The only difference is the use of a weighted regional utility variable in place of the discrete regional utility variable used in the discrete nested logit model. A destination's weighted regional utility is not simply the summation of the utility values of itself and all the other districts within the same discrete region, but instead is the weighted sum of ALL districts' utilities. The contribution of any particular district's utility to the weighted regional utility of destination A is weighted by the likelihood of a migrant cognizing that destination in the same region as destination A, as indicated by the appropriate cell of the probabilistic regionalization matrix. The weighted regional utility variable is formulated in equation 5.9 below. This formulation can be compared with the discrete regional utility formulation shown in equation 5.7 above.

$$WR_j = \sum_k W_{jk} \, (D_{k1}^{\lambda 1} \, D_{k2}^{\lambda 2} \ldots D_{kg}^{\lambda g}) \qquad \textit{(Eq. 5.9)}$$

Where:

WR_j is the weighted regional utility of destination j
W_{jk} is the weighting of district k's utility (i.e. element (j,k) in regionalization matrix)
$D_{k1}^{\lambda 1}$ is parameterised explanatory variable 1 describing destination k

Equation 5.9: Weighted regional utility.

The formulation in equation 5.9 gives rise to a unique weighted regional utility value for each and every destination. This contrasts with the discrete regional utility value of the discrete nested logit model, which has the same value for all destinations allocated to the same discrete region. It should also be stressed that in order to effectively capture regional utility variation, the calculation of each destination's weighted regional utility involves a weighted summation of ALL the other destinations' individual utilities – not just the utilities of the 100 selected districts in the migration sub-system under consideration in this research.

Operationally, the weighted nested logit model is calibrated in a very similar manner to the discrete nested logit model as two sequential logit calibrations. The first logit model calibration is essentially a traditional logit model calibration. The parameter estimates resulting from this stage-one calibration are then used to calculate values of the weighted 'region utility' variable for each potential destination using the weighted summation mechanism described above. In this context the term 'regional' utility could be considered misleading, given that every potential migration destination will very likely have a unique weighted regional utility value due to the way this variable is calculated. The 'regional utility' terminology is used deliberately here in order to highlight the parallels with the discrete regional utility variable of the discrete nested logit model. In both nested logit models the regional utility variable is intended to capture the hierarchical structure of the destination choice set which this thesis proposes is fundamental to the migration destination selection process.

Assuming that migration destination selection is indeed the spatially hierarchical process proposed above, such that consideration of regional attributes precedes examination of individual destination characteristics, then the regional utility of a destination can also be considered to indirectly represent the likelihood of a destination even appearing in a migrant's destination choice set at the final stage of their choice process when they are comparing individual destinations. In this way, the effect of the nested logit models' regional utility variables can be considered to be somewhat similar to that of the competing destinations model's accessibility variable – all are choice set modifiers.

The weighted nested logit model has some parallels with the geographically weighted regression (GWR) approach to spatial modelling developed by Brunsden, Fotheringham and Charlton insofar as the weighted manner in which the regional utility variable is calculated is somewhat similar to the spatial kernel weighting of explanatory variables in GWR, which can apply complex distance decay functions to vary the spatial scale over which different variables values will be weighted when calibrating the regression model (Brundsen, Fotheringham and Charlton, 1998; Fotheringham, Brunsden and Charlton, 2000).

Hybrid weighted nested logit model

Though there are parallels in causal interpretation between the regional utility variables of the nested logit models and the accessibility variable of the competing destinations model, these variables are derived from different theoretical perspectives. There is also empirical evidence, that will be presented in detail in chapters 7 through 10, that there are fundamental spatial and, to a lesser extent, socio-economic differences in the mode of operation of these variables. These theoretical and empirical differences, and the minimal correlation observed between the variables, suggested that both variables might usefully coexist in a single model that could potentially combine the predictive benefits of both the competing destinations and nested logit approaches to provide better predictive ability than either model alone.

A new model, incorporating both the accessibility variable from the competing destinations model and the weighted regional utility variable from the weighted nested logit model was derived and termed the hybrid weighted nested logit model. The formulation for this model is presented in equation 5.10 below.

$$M_{ij} = M_i \frac{exp\ (D_{j1}^{\lambda 1}\ D_{j2}^{\lambda 2}\ ...D_{jg}^{\lambda g}\ A_j\ WR_j^{a})}{\sum_k exp\ (D_{k1}^{\lambda 1}\ D_{k2}^{\lambda 2}\ ...D_{kg}^{\lambda g}\ A_k\ WR_k^{a})} \qquad (Eq.\ 5.10)$$

Where:

M_{ij} is the migration from origin i to destination j
M_i is the total out-migration from origin i
$D_{j1}^{\lambda 1}$ is parameterised explanatory variable 1 describing destination j
A_j is the accessibility of destination j
WR_j^{a} is parameterised weighted regional utility for destination j

Equation 5.10: The hybrid weighted nested logit model.

Despite the differing theoretical derivation of the four hierarchical migration destination choice models presented in this chapter, all share operational similarities which mean that the same software can be adapted to calibrate all of these models.

Model calibration

All the models employed in this analysis: traditional, competing destinations, discrete/weighted nested logit and the hybrid model, can be considered, in operational terms, to be share logit models, or pairs of sequential share logit models. For instance, whilst the accessibility variable of the competing destinations model acts as a choice set modifier in theoretical terms, in operational terms it is just another explanatory variable in a logit share model. The nested logit models are calibrated as two sequential share logit models: first a traditional share-logit model is calibrated (or in the case of the hybrid model, an initial competing destinations model is calibrated), then a second-stage share-logit is calibrated that contains an additional explanatory variable representing each destination's regional utility, as calculated using the initial calibration's parameter estimates.

The dependent variable of the share-logit migration models applied here are counts of migrants selecting each of a number of possible alternative destinations, thus they are Poisson regression models. All the Poisson regression models employed in this analysis were calibrated using a spatial interaction modelling software package called SIModel written in the early 1980s by Williams and Fotheringham (Williams and Fotheringham, 1984). SIModel is capable of calibrating a wide range of spatial interaction models. For the purposes of this analysis SIModel was used to calibrate doubly-constrained global logit-share models and production-constrained origin-specific logit-share models. SIModel produces maximum likelihood estimates of the models' parameters and outputs them along with their standard errors, observed-predicted flows and a host of goodness-of-fit and related statistics.

The calibration of batches of simple and step-wise, global and origin-specific models was automated, for a variety of migrants subgroups, using simple programs written in C – that called onto underlying operating system commands to move files around and invoke commands. Such scripts were also used to tabulate and efficiently consolidated the multitude of data output by the SIModel software, as well as to generate the accessibility variables used by the competing destinations model. The source code of these model automation programs is included as appendix C to this thesis to facilitate reproduction of the results contained herein.

Summary

This chapter has completed the history of spatial interaction modelling begun in chapter 2, showing how modern migration destination choice models can be considered to be the fourth-generation of spatial interaction models that have been applied to migration research. First came the gravity models of Ravenstein based on his principles of social physics. These were followed by Wilson's models derived from principles of maximizing entropy. Then logit-based discrete choice modelling from aspatial applications, particularly in economics, constituted the third generation of models applied to migration destination choice. Finally, we now have the incorporation of the effects of spatial structure into the logit framework in order to model the likelihood of any particular destination appearing in any particular migrant's perceived choice set.

This chapter has introduced four fourth-generation migration models that use various mechanisms to represent how spatial structure affects the likelihood of any particular destination being considered in any particular migrant's destination choice process, and thus affects migration destination choice behaviour.

The competing destinations model includes a destination accessibility variable to represent each destination's likelihood of being cognized in a cluster along with other potential destinations. Significant estimates for this parameter would suggest that migrants are indeed mentally clustering destinations and considering some destination characteristics at a cluster-level rather than independently for each individual destination. Steven's 'psychophysical law' that the size of a cluster is increasingly underestimated as actual cluster size increases, suggests that the accessibility parameter estimates will generally be negative, implying that destinations that are more likely to be cognized in a cluster with other destinations are *ceteris paribus* less likely to be selected as migration destinations.

The discrete and weighted nested logit models assess the likelihood of any particular destination being considered by any particular migrant by including a variable representing the utility of each destination's containing region, it being less likely for an individual destination to be selected if its containing region has a lower utility compared to others. When calibrating discrete nested logit models the regions are defined discretely with every destination being in one and only one region. In the case of the weighted nested logit model the regions are defined probabilistically, with a set of likelihood values representing the likelihood of any particular destination being cognized together. The 'weighted regional utility' associated with any particular destination is the sum of all other destinations' utilities weighted by their likelihood of being cognized in the same destination cluster as the destination under consideration. Again, significant parameter estimates for the 'regional utility' variables in the discrete and weighted nested logit models are indicative of a hierarchical destination choice process in which regional characteristics determine which region's destinations are then considered in more detail.

The hybrid weighted nested logit model combines the benefits of accessibility and regional utility variables from the competing destinations and weighted nested logit models, respectively, into a single hybrid migration destination choice model.

Whilst the discussion above has introduced the manner in which the competing destinations, nested logit and hybrid models of migration destination choice incorporate the notion of destination clustering and the importance of regional characteristics, it has not discussed how these clusters and regions are defined in practice. The next chapter discusses how such regionalizations, or hierarchical destination choice sets, can be generated.

Chapter Six

Defining Hierarchical Migration Choice Sets

Introduction

Central to any hierarchical model of migration destination choice is the definition of an appropriate hierarchical choice set. The competing destinations model does not require a discrete hierarchical choice set, with each destination allocated to one and only one region. Instead it uses an accessibility variable to represent the likelihood of a particular destination being cognised by a migrant as part of a larger cluster of destinations. The nested logit models applied here are calibrated within the context of both discrete and weighted origin-specific regionalizations, which are intended to directly represent the manner in which migrants group potential destinations. Details are presented below of how to generate these various types of hierarchical choice sets.

Competing destinations choice sets

As described above, calibration of the competing destinations model does not require an explicitly defined hierarchical choice set. Instead, the model incorporates the likelihood of each potential destination being cognised within a larger group of potential destinations. This is achieved by including in the model an explanatory variable representing the 'accessibility' of each destination district. This accessibility statistic represents how many other destinations are in close proximity to it. Put in a more cognitive context, the accessibility variable represents the likelihood that a particular destination will, based solely on its location relative to other potential destinations, be considered as a distinct destination by a migrant, rather than being combined with other neighbouring destinations into a cluster, or being ignored/overlooked altogether due to larger, better-known neighbours. In other words it represents the degree of competition that each destination experiences from other destinations for the attention of, and therefore consideration by, each migrant – hence the name of this model.

Including an accessibility variable in the model accounts for the effect upon destination choice of the relative spatial situations of all potential destinations, and thus the competing destinations model does not suffer from the Independence from Irrelevant Alternatives (IIA) property which the traditional logit model suffers as discussed in the previous chapter. In the competing destinations model the alternatives are very relevant indeed – as it is their size and relative location that determines the value of the accessibility variable that is central to the model.

The inclusion of an accessibility statistic is based upon the assumption that migrants select migration destinations hierarchically, meaning that they will mentally partition the potential destinations into groups. We have seen that perception of group size is non-linear, such that the size of groups or clusters is increasingly underestimated as they increase in size, (Stevens, 1975). As stated in the previous chapter, from this finding one would anticipate that a region of 20 destinations will be perceived to offer less than twice the utility of a region made up of 10 'identical' destinations. Furthermore, this underestimation of the larger region's aggregate utility will cause migrants to select it over the smaller region less often than its actual size would justify.

The upshot of this is that an individual destination within a larger group of destinations is less likely *ceteris paribus* to be selected than a destination in a smaller group of destinations. Thus, because accessibility represents the likelihood of a destination being cognised within a larger cluster of destinations, the calibration of a competing destinations model will yield a negative estimate for the accessibility variable parameter if such hierarchical decision making behaviour is indeed taking place.

It is by no means self-evident exactly how such an accessibility statistic should be derived. Plane and Mulligan (1997) suggested that a useful measure of the spatial structure in a migration system can be derived directly from examining the observed migration within that system. Specifically, they proposed that an area's 'accessibility' can be related to the characteristics of the arrays of in- and out-migrations to and from all other areas, to get an indication of how 'well-connected' that area is. However, there is no clear psychological basis to support this approach to capturing spatial structure, so we are here following the more widely accepted assumptions that the two key components of accessibility are the size and locations of competing destinations relative to each other. It is intuitive to expect that destinations that are closer together are more likely *ceteris paribus* to be considered in the same destination cluster by a migrant. It is also reasonable to expect that destination clusters will be centred about larger areas as these are the areas about which migrants generally have the most information. The simplest definition of an accessibility statistic that takes into account these two factors can be expressed as:

$$a_i = \sum_{\substack{j=1 \\ j\neq i}}^{N} (p_j / d_{ij}) \qquad \textit{(Eq. 6.1)}$$

Where:

N is the total number of destinations
p_j is the population of destination j
d_{ij} is the separation of destinations i and j.

Equation 6.1: Simple formulation of accessibility statistic.

The formulation presented in equation 6.1 may also be familiar as the definition of the population potential of area i. Values of population potential are most often used in a comparative measure of centrality and as such absolute values of population potential are not usually of critical importance. Therefore the implicit use of +1.0 and –1.0 as the population and distance exponents, respectively, when calculating population potential is a perfectly reasonable simplification.

However, when calculating an accessibility statistic to represent the competition due to relative size and spatial location, there is no reason to assume such linear relationships between the size and separation and the degree of competition that a destination will experience. Thus, equation 6.1 is extended by the addition of population and distance exponents:

$$a_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} (p_j^{\alpha} / d_{ij}^{\beta}) \qquad \textit{(Eq. 6.2)}$$

Where:

N is the total number of destinations
p_j^{α} *is the parameterized population of destination j*
d_{ij}^{β} *is the parameterized separation of destinations i and j.*

Equation 6.2: Parameterized formulation of accessibility statistic.

The selection of appropriate values for these population and distance exponents is an area of some debate (Fotheringham 1983, 1991). In theory one could iteratively derive the population and distance exponent values by repeatedly calibrating a competing destinations model, each time regenerating the accessibility variable using the distance and population parameter estimates from the preceding model calibration, (having set them arbitrarily to, say, 1.0 and –1.0 to calculate the accessibility statistic for the initial run of the model). However, there is no clear theoretical basis for deriving the accessibility exponents in this manner, and one cannot be certain that this method would actually converge upon a single set of values. Furthermore, one could argue that this approach is flawed because the parameter estimates generated by calibrating the model are primarily representing the population pull and distance deterrence rather than the competition between the destinations – though in the absence of an accessibility variable they are also affected by the relative positioning of the various destinations. Obviously the introduction of the accessibility variable is intended to remove this component from the population and distance parameter estimates, so that they become more accurate reflections of the population pull and distance deterrence effects.

It is not unreasonable to suggest, however, that the scale at which the accessibility variable is calculated should be appropriate to the scale at which the migration destination choice process is being modelled. For instance, if the distance exponent in the accessibility equation is too low, then for any population exponent, the variation in the value of the accessibility variable will be much more gradual across space. Thus the level of differentiation that the accessibility variable provides between destinations a given distance apart will be reduced.

In order to determine suitable distance and population exponents to be used in the migration destination choice modelling undertaken in this research, 144 different accessibility variables were calculated using a range of distance and population exponent combinations, (from 0.25 up to 3.0 in 0.25 increments, for both exponents). Each of these accessibility variables was then used to calibrate a set of competing destinations models and the results were used to rank the 144 accessibility variables according to goodness-of-fit and the significance of the accessibility parameter estimates they produced.

More specifically, for each of the 144 accessibility variables generated, 100 origin-specific competing destinations models were calibrated. The average R^2_{adj} goodness-of-fit statistic from these 100 model calibrations was calculated along with a count of how many of the 100 models' accessibility parameter estimates were statistically significant at the 99% level. The higher than usual significance level of 99% was used in order to better differentiate the model runs, because at 95% confidence most origins' accessibility parameter estimates were significant for most of the 144 accessibility variables. When these two sets of results, the average R^2_{adj} statistics and the count of significant parameter estimates, had been collected for each of the 144 sets of model runs, the accessibility variables were ranked from highest to lowest average R^2_{adj} and most to least significant accessibility parameter estimates. Tables 6.1 and 6.2, below, present these rankings, tabulating the distance and population exponents used to create the accessibility variables and showing how the resulting accessibility variables ranked against each other.

	D0.25	D0.50	D0.75	D1.00	D1.25	D1.50	D1.75	D2.00	D2.25	D2.50	D2.75	D3.00
P0.25	108	102	92	90	99	111	121	128	135	141	143	144
P0.50	105	94	84	78	87	100	114	123	131	137	140	142
P0.75	101	86	74	71	72	79	103	115	125	132	136	138
P1.00	95	75	64	57	55	65	77	104	116	127	130	134
P1.25	89	69	52	47	43	48	59	81	107	120	126	129
P1.50	83	60	46	34	30	33	42	61	88	109	118	124
P1.75	85	56	37	26	19	20	28	44	66	91	110	119
P2.00	96	58	36	21	13	11	17	29	49	73	97	112
P2.25	113	67	39	23	10	2	7	18	35	53	76	98
P2.50	122	82	51	31	16	4	1	8	24	41	62	80
P2.75	133	106	70	45	27	14	5	3	15	32	50	68
P3.00	139	117	93	63	40	25	12	6	9	22	38	54

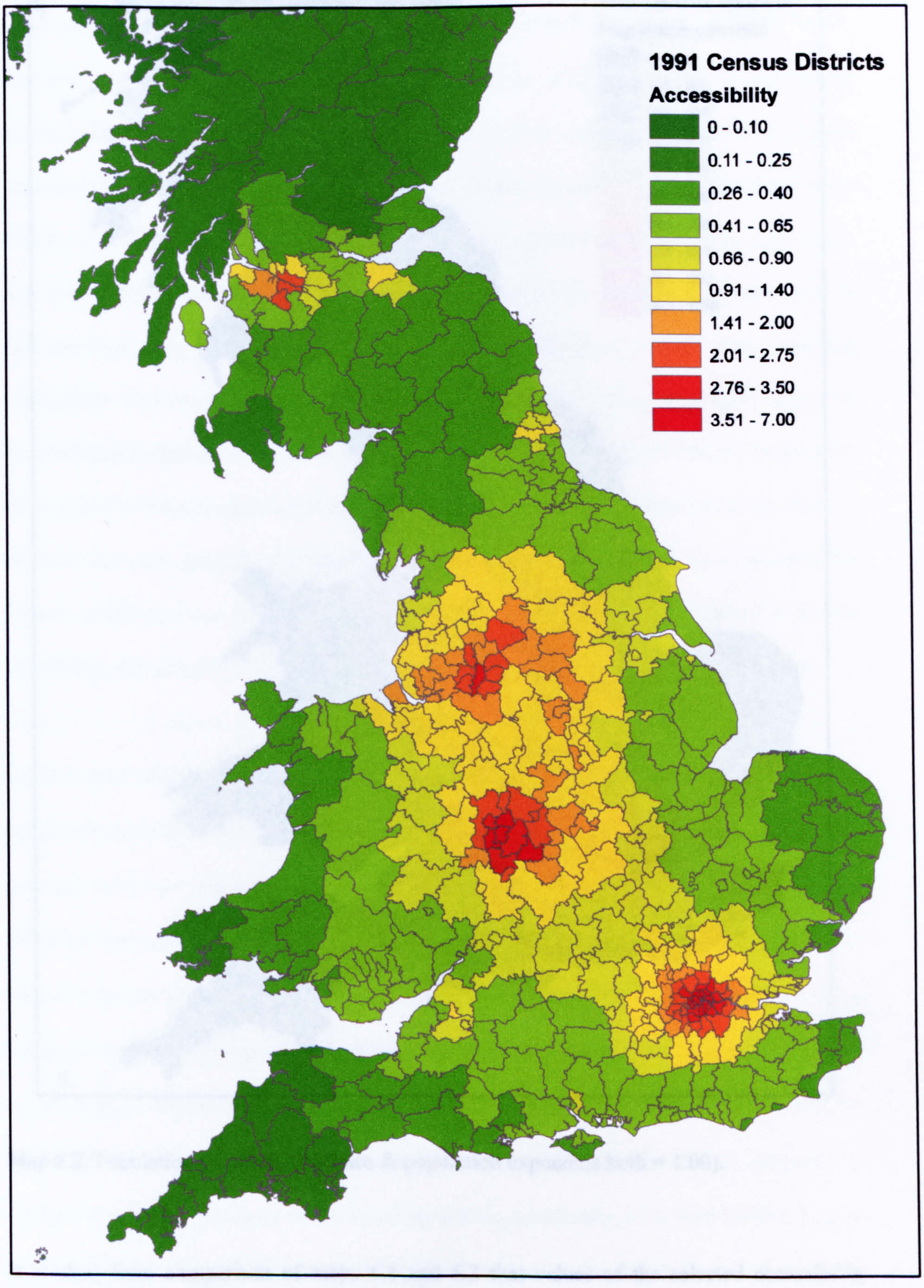
Table 6.1: Accessibility variables ranked by average origin-specific R^2_{adj} goodness-of-fit.

	D0.25	D0.50	D0.75	D1.00	D1.25	D1.50	D1.75	D2.00	D2.25	D2.50	D2.75	D3.00
P0.25	122	122	122	122	109	116	122	129	135	141	143	144
P0.50	109	109	116	109	104	116	116	104	129	136	139	142
P0.75	94	109	104	100	100	94	94	94	116	132	136	140
P1.00	91	104	100	91	82	85	85	85	91	109	132	136
P1.25	82	82	72	57	57	57	57	57	77	94	127	132
P1.50	43	43	36	43	36	43	43	43	57	72	109	129
P1.75	43	36	29	29	19	36	36	29	19	43	77	116
P2.00	43	19	19	6	6	6	12	19	29	57	85	100
P2.25	77	29	1	1	1	1	12	19	43	72	57	94
P2.50	85	43	6	1	6	12	12	36	57	57	72	77
P2.75	104	77	12	6	12	12	19	36	43	57	57	72
P3.00	127	85	57	19	19	19	29	29	43	43	57	57

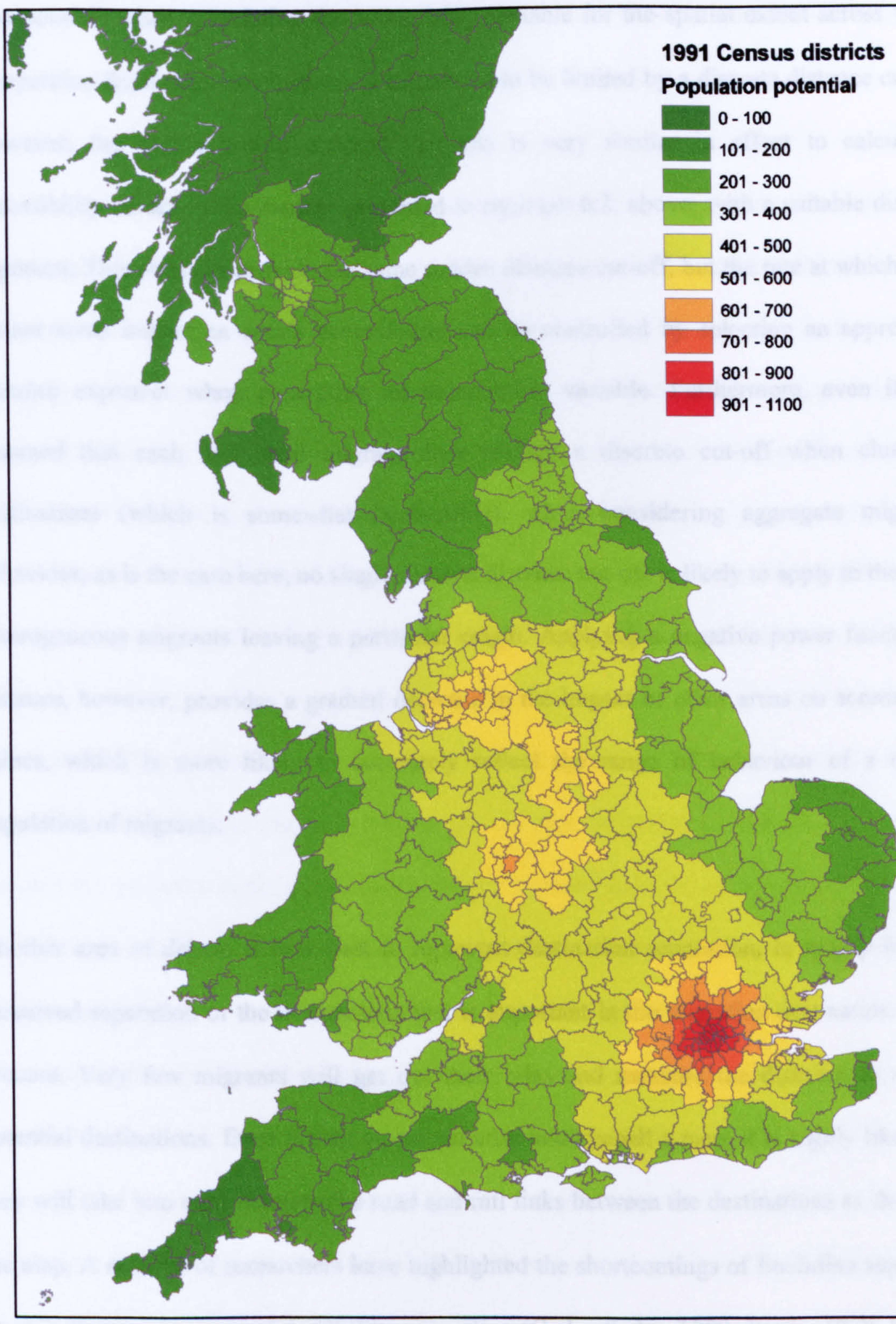
Table 6.2: Accessibility variables ranked by accessibility parameter estimate significance.

Note that the parameter estimate significance ranking includes some equal rankings for those accessibility variables that resulted in an equal number of the 100 origin-specific calibrations having 99% significant accessibility parameter estimates.

Whilst these two attempts to determine optimal population and distance exponent values accessibility variable generation do not concur precisely, they do identify similar ranges of exponent values as performing best. Based on these results the values of 2.50 and 1.50 were selected for the population and distance exponents, respectively, and the accessibility variable generated from these exponent values was considered the definitive accessibility variable for use in all other competing destinations and hybrid model calibrations throughout this research. The accessibility values obtained from use of these exponent values is presented in map 6.1, below. For comparative purposes, a map of population potential is shown in map 6.2 – recall that this is essentially an accessibility variable produced using population and distance exponents of 1.0.



Map 6.1: Accessibility based on distance and population exponents of 1.50 & 2.50.



Map 6.2: Population potential, (distance & population exponents both = 1.00).

It is clear from comparison of maps 6.1 and 6.2 that values of the selected accessibility variable vary more rapidly over space and are influenced more by major population centres, than are values of population potential.

It is possible when calculating the accessibility variable for the spatial extent across which competition from other destinations is considered to be limited by a discrete distance cut-off. However, the author would contend that this is very similar in effect to calculating accessibility values in the manner presented in equation 6.2, above, with a suitable distance exponent. This does not provide the same sudden distance cut-off, but the rate at which more distant areas impact an area's accessibility can be controlled by selecting an appropriate distance exponent when generating an accessibility variable. Furthermore, even if it is assumed that each individual migrant does employ a discrete cut-off when clustering destinations (which is somewhat implausible), when considering aggregate migration behaviour, as is the case here, no single discrete distance cut-off is likely to apply to the many heterogeneous migrants leaving a particular origin. Applying a negative power function of distance, however, provides a gradual decrease in the impact of other areas on accessibility values, which is more likely to accurately reflect the range of behaviour of a diverse population of migrants.

Another area of debate is how best to represent destination separation. In reality it is the perceived separation of the destinations that is important in the migration destination choice process. Very few migrants will get out their atlas and measure the distance to various potential destinations. Even if they do use an atlas and eyeball a map, it is highly likely that they will take into consideration the road and rail links between the destinations as shown on the map. A number of researchers have highlighted the shortcomings of Euclidian separation as an indicator of perceived destination separation (Cadwallader, 1979; Evans, 1980; Sadella, 1980). Inaccuracies of Euclidian place separation are often due to natural geographical factors, such as the geometry of the coastline and the positioning of natural barriers such as rivers, mountain ranges or barren moorland. For instance, in Great Britain, consider the case of Cardiff and Exeter - the straight-line Euclidian distance between these cities is unlikely to be a good representation of migrants' mental separation of the two destinations because of the geographical barrier of the River Severn estuary.

One alternative which has been proposed is to base the separation variable upon typical road and rail travel times between destinations. However, despite the availability of GIS tools and computational network analysis techniques which can simplify the generation of such data, the majority of migration analysis continues to make use of simple Euclidian straight-line distances as a surrogate for migrants' perception of destination separation, and for simplicity, the same variable is used in this research. If such travel-time or survey-derived place separation data were available it would be interesting to examine how its use affected the distance variable's parameter estimates from the calibration of these models.

The accessibility variable contains no information about linkages between specific destinations and therefore does not define a specific regionalisation of destinations. However, because the inclusion of the accessibility statistic is based upon the assumption that migrants mentally group their potential destinations, so it can be said the competing destinations model assumes a weighted regionalisation of destination districts even though the model is not explicitly calibrated within the context of any regionalization. This is a particularly beneficial characteristic when this model is calibrated against aggregate migration data, as in this research, since no single discrete regionalisation could hope to accurately represent the multitude of mental hierarchies that individual migrants would no doubt employ when deciding upon their migration destination. This is a criticism that is often levelled at the nested logit model when it is applied to the analysis of migration. The nested logit family of models, and the generation of the regionalizations within which they are calibrated, are discussed in the following section, which also introduces a novel technique for generating and incorporating probabilistic regionalizations into the nested logit modelling framework.

Nested logit choice sets

The previous chapter introduced several nested logit models and described how their calibration requires a pre-defined choice hierarchy.

The discrete nested logit model requires a mutually exclusive discrete allocation of destinations to regions. The research reported here applies two-level discrete nested logit migration models; thus the hierarchical choice sets are simply allocations of all destinations to regions, with the region and individual destination being the two spatial scales at which explanatory variables are considered. A novel approach to the generation of such discrete regionalizations was developed by the author and is presented below.

Discrete regionalisation generation

It is theoretically possible, though far from trivial, to employ deterministic agglomerative techniques to generate an optimal discrete regionalisation. However, the additional complexity of such an approach was not considered worthwhile, or even desirable, in this research. Instead, a novel regionalisation algorithm was employed that introduces some random variation into the regionalization process. This was felt to be more appropriate for two reasons:

- No single discrete regionalization, whether generated deterministically or otherwise, can effectively capture and represent the inevitable random variation between the mental hierarchies of the many heterogeneous migrants leaving a specific origin.
- It is non-trivial to capture the thought-processes that an individual migrant employs when allocating destinations to a spatial hierarchy and more difficult still to map these to algorithmic heuristics that could drive deterministic agglomerative choice set generation. This is exacerbated by the fact that many aspects of such mental destination clustering are very likely to be subconscious, and will inevitably vary between individual migrants.

The regionalization algorithm employed here is based upon the assumption that migrants will create regions containing broadly similar amounts of information about their constituent destinations. This is intuitively plausible, as one would expect migrants to be able to differentiate to a greater extent between destinations about which they have more information, and therefore are more likely to be able to create more specific groupings of such destinations. This assumption also forms the basis of the measure of regionalisation quality that is used to assess and rank the regionalizations generated for a specific origin. Specifically, regional information variance is the characteristic by which regionalizations are compared, with smaller values representing 'better' regionalisations.

Also, in the same way that migrants are assumed to be able to better differentiate between destinations with which they are more familiar, it is also assumed that migrants who are generally better informed, because they live in a more central or accessible location, will be able to differentiate destinations in general to a greater extent. This factor is operationalized by creating regionalisations comprising a larger number of constituent regions for migrants leaving more central origins.

The algorithm by which discrete regionalisations are generated for use in this research is described below. The full source code of this regionalization algorithm can be found in appendix B. Note that each regionalization is generated separately for migrants leaving each origin. Also, the regionalizations allocate every district to a region, not just the 100 study areas selected for this research, because the values of the regional utility variable used in the discrete nested logit model are influenced by the utilities of all 459 districts.

The discrete regionalization algorithm is summarized as a flow chart in figure 6.1, below, followed by a more detailed description of each step of the process.

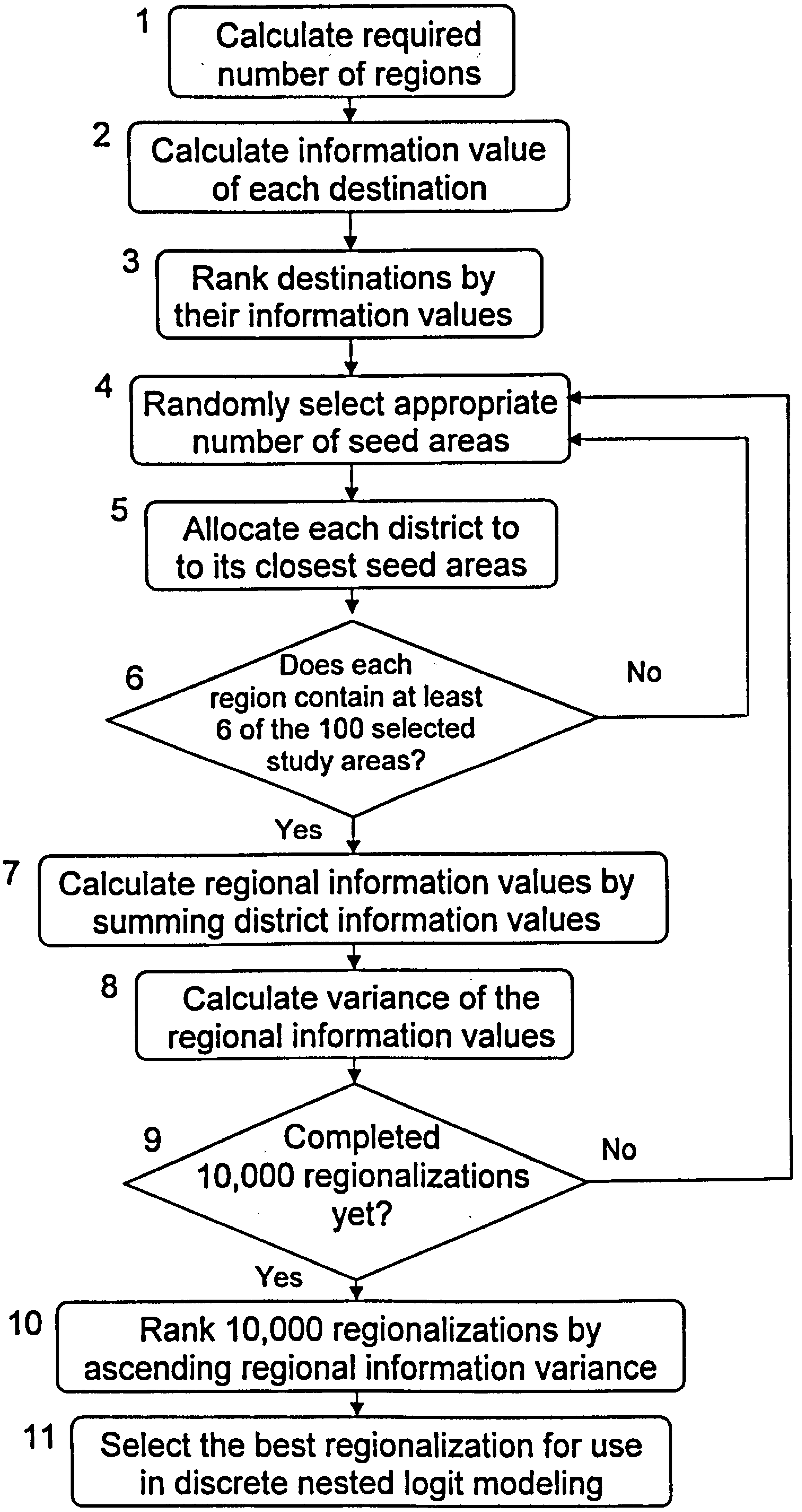


Figure 6.1: Flow chart of discrete regionalization algorithm.

Step 1: Calculate the required number of regions

It has been proposed above that the complexity of a migrant's destination choice hierarchy is likely to vary between migration origins. It is not unreasonable to expect that migrants from a more central origin will generally have information about more potential destinations, than a migrant from a more remote origin. If migrants from more central origins have more destination information then they will *ceteris paribus* be better able to differentiate between destinations and will therefore be likely to perceive smaller groups or clusters that they consider to contain similar destinations. When generating discrete regionalizations this variation in choice set complexity between migrants from different origins is operationalized by relating the number of regions in each regionalization to the centrality, or population potential, of the origin under consideration. It will be recalled from equation 6.1 that an origin's population potential is the sum of each other destination's population divided by it's separation from the origin under consideration.

The precise relationship that was chosen between an origin's population potential and the number of regions to be generated in discrete regionalizations for that origin is influenced by the desire to maintain reasonable numbers of choice alternatives both between and within regions, whilst also representing the notion that migrants from more central origins are likely to construct more complex mental maps. Specifically, a constraint was imposed that there should be at least six alternatives at each level of the hierarchical choice set. This same constraint also dictates the upper limit on the number of regions to be generated - the higher the number of regions, the less likely it is that this can be achieved. In order for the regionalization algorithm to generate a suitable number of discrete regionalizations within a reasonable period of time, it was found that an upper limit of nine regions in any regionalization was effective. Beyond that number, the regionalization process could take over a week to produce sufficient regionalizations for some origins. The actual formula used to calculate the number of regions to product is presented in equation 6.3 below:

$$N_i = int(5.5 + \frac{4 * (P_i - P_{min})}{(P_{max} - P_{min})}) \qquad (Eq.\ 6.3)$$

Where:

N_i is number of regions to be generated for migrants from origin i
int() is a function that rounds to the nearest integer
P_i is population potential of origin i
P_{max} is the maximum population potential of any origin
P_{min} is the minimum population potential of any origin

Equation 6.3: Formula for the number of regions required.

The formula in equation 6.3 results in values between 6 and 9 for all origins except the origin with the highest population potential for which an N_i value of 10 results (which is reduced to 9 by a special case in the regionalization computer program). N_i for the area with the minimum population potential equates to:

$$N_{Pmin} = int(5.5 + (4 * 0)/(P_{max}-P_{min})) = int(5.5) = 6.$$

The number of regions that equation 6.3 prescribes for the origin with the highest population potential is:

$$N_{Pmax} = int(5.5 + (4*(P_{max}-P_{min})/(P_{max}-P_{min}))) = int(5.5 + 4) = int(9.5) = 10$$

Thus, the application of equation 6.3 to the 100 selected migration origins produces a quantitative representation of the complexity of the regionalization that should be generated for use when modelling migration from each origin, or more specifically, the number of regions that each origin's regionalizations must contain.

Step 2: Calculate the information value of each district

The basis of the discrete regionalization process is that migrants will be able to differentiate better between those destinations about which they have more information, and will therefore aggregate those destinations about which s/he is better informed into smaller clusters or regions. In this research the amount of information that a migrant has about each destination is calculated by a very simple formula:

$$I_{ij} = p_j / d_{ij} \qquad \text{(Eq. 6.4)}$$

Where:

p_j *is the population of district j*

d_{ij} *is the separation of districts i and j*

Equation 6.4: Information that migrant from origin i has about district j.

Using the simple formula presented in equation 6.4 the information that a migrant from the origin under consideration has about each of the other 458 districts is calculated. Note that all 458 districts are considered as potential destinations when generating discrete regionalizations, so information values are calculated and ranked for all 458 districts (other than the origin district). This is because the regional variables that are central to the operation of the nested logit models are derived from the utility values of all districts within each region, not just those that are also amongst the 100 selected study areas.

Step 3: Rank districts according to their information values

The next step of the regionalization process is to order the 458 destinations according to the amount of information that migrants from the origin under consideration are calculated to have about each destination, ranking from most information to least.

Step 4: Randomly select an appropriate number of seed areas

Once the 458 districts have been sorted into 'information order', an appropriate number of seed areas are selected from these 458 districts to form the focal points of the discrete regions. The seed areas are selected from specific quartiles of the information-ranked list of districts according to the distribution shown in table 6.3 below.

Number of regions required:	Quartile 1	Quartile 2	Quartile 3	Quartile 4
6	2	2	1	1
7	3	2	1	1
8	3	2	2	1
9	3	3	2	1

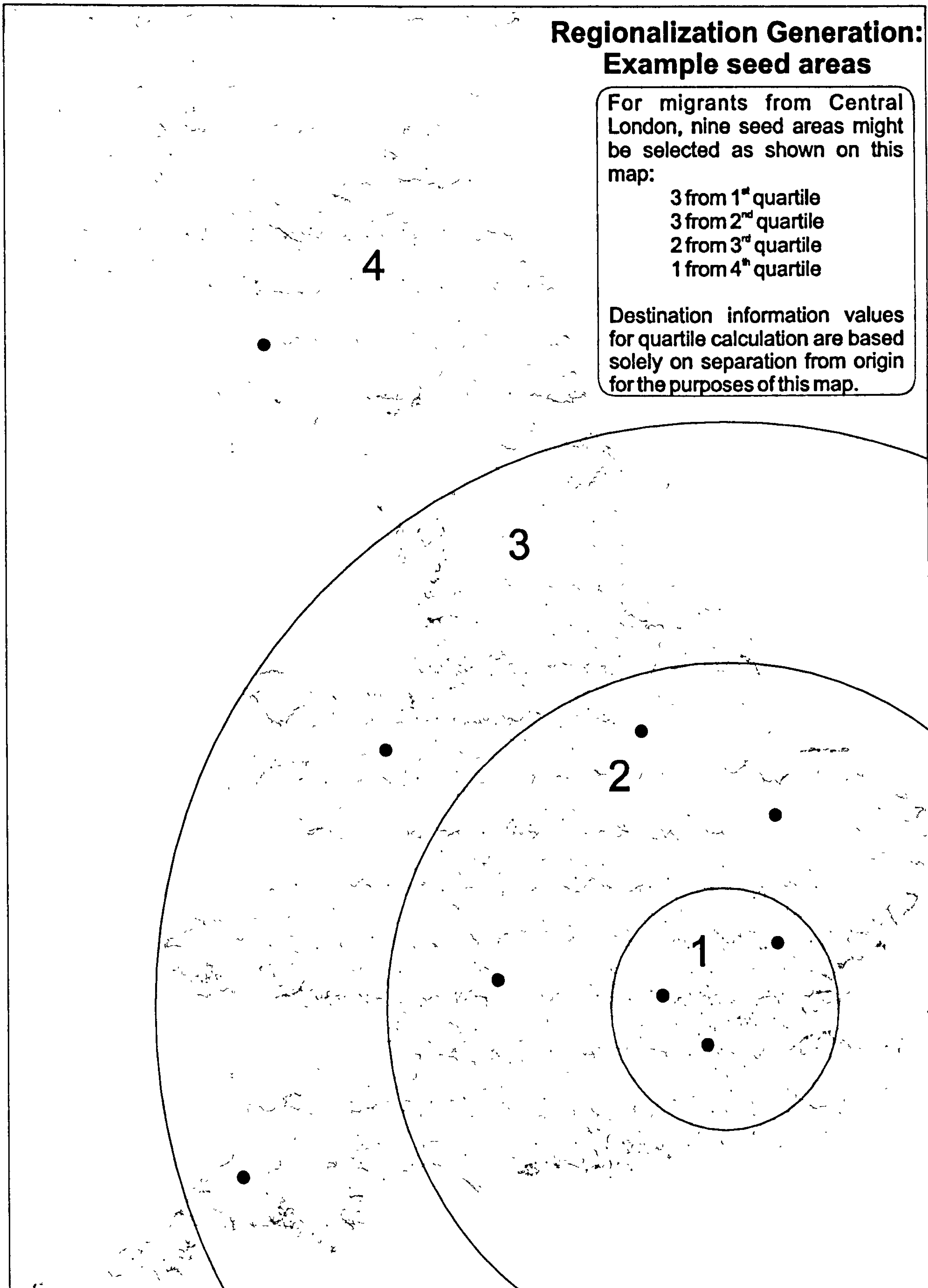
Table 6.3: Per-quartile allocation of seed areas for different regionalization complexities.

This means that when generating a regionalization comprising eight regions, three seed areas will be selected from the districts in the top quartile of the information-ranked district list, two from the second quartile, two from the third quartile and one from the quartile of districts about which a migrant from the origin in question has the least information.

These quartiles correspond with districts, 1-115, 116-230, 231-345 and 346-459 in the information-ranked district list, so appropriate random numbers in each of these ranges are selected to form the basis of the required number of regions. This approach ensures that there will be more and smaller regions in those areas that a migrant from an origin has more information about – i.e. typically areas closer to that origin.

This process is clarified by the simplified illustration in map 6.3, below. For the puposes of this explanatory map destination information is calculated on the basis of separation from origin only, thus the first quartile of destinations is the 115 closest to the origin, i.e. those within the inner circle. With the fourth quartile being made up of: the tip of Cornwall; far Northern England; and Scotland - all of which lie outside the outer circle. Map 6.3 shows an example of how nine seed areas might be selected from these four quartiles: three each from the first and second quartiles, two from the third quartile and one from the fourth (as per table 6.3 above).

Map 6.3: Example seed area selection from information-ranked destination quartiles.



Step 5: Form regions by allocating each district to its closest seed area

The seed areas selected in step 4, above, represent the foci of the regions. The regions are discretely defined by now associating each and every district with its closest seed area.

Step 6: Check regionalization meets constraints required for step-wise calibration

In order for a regionalization to be considered valid, the constraints of the step-wise nested logit calibration process dictate that each region must contain at least five choice alternatives. This number was increased to six for the purposes of this research to avoid working at the margins of the calibration mechanism's applicability. The choice alternatives in this context correspond with the 100 selected study areas, not to districts in general – thus, every region of the regionalization must contain at least six selected study areas for it to be considered further in this analysis.

Step 7: Calculate regional information values

Because there is a random element in the selection of the seed areas, it is inevitable that many intuitively unacceptable regionalizations will be produced by this process – regionalizations that it is very unlikely represent the mental clustering of destinations of many, if any, migrants. It is likely that some such unlikely regionalizations may even meet the constraint checked in step 6, above, but it is desirable that these regionalizations nonetheless be filtered out and not employed in this analysis. The mechanism chosen to assess the quality of the regionalizations is the variance in regional information values. The theoretical basis for this discrete regionalization process, presented above, suggests that it should produce more and smaller regions in those areas about which migrants from the origin under consideration have the most information. This will have the effect of balancing regional information values, so those regionalizations exhibiting lower regional information variance are the ones that best represent the intent of this regionalization process.

Regional information values are calculated by summing the information values of each region's constituent districts, from the perspective of migrants from the origin under consideration.

Step 8: Calculate regional information variance

From the regional information values calculated in step 7 the regional information variance is calculated for each regionalization.

Step 9: Repeat steps 4-8 10,000 times

Because the regionalization generation process is partly random, quality of regionalizations is assured by generating a large number of regionalizations, ranking them by a quality measure and then selecting the best. In this research 10,000 valid regionalizations were generated relative to migration from each of the 100 selected migration origins.

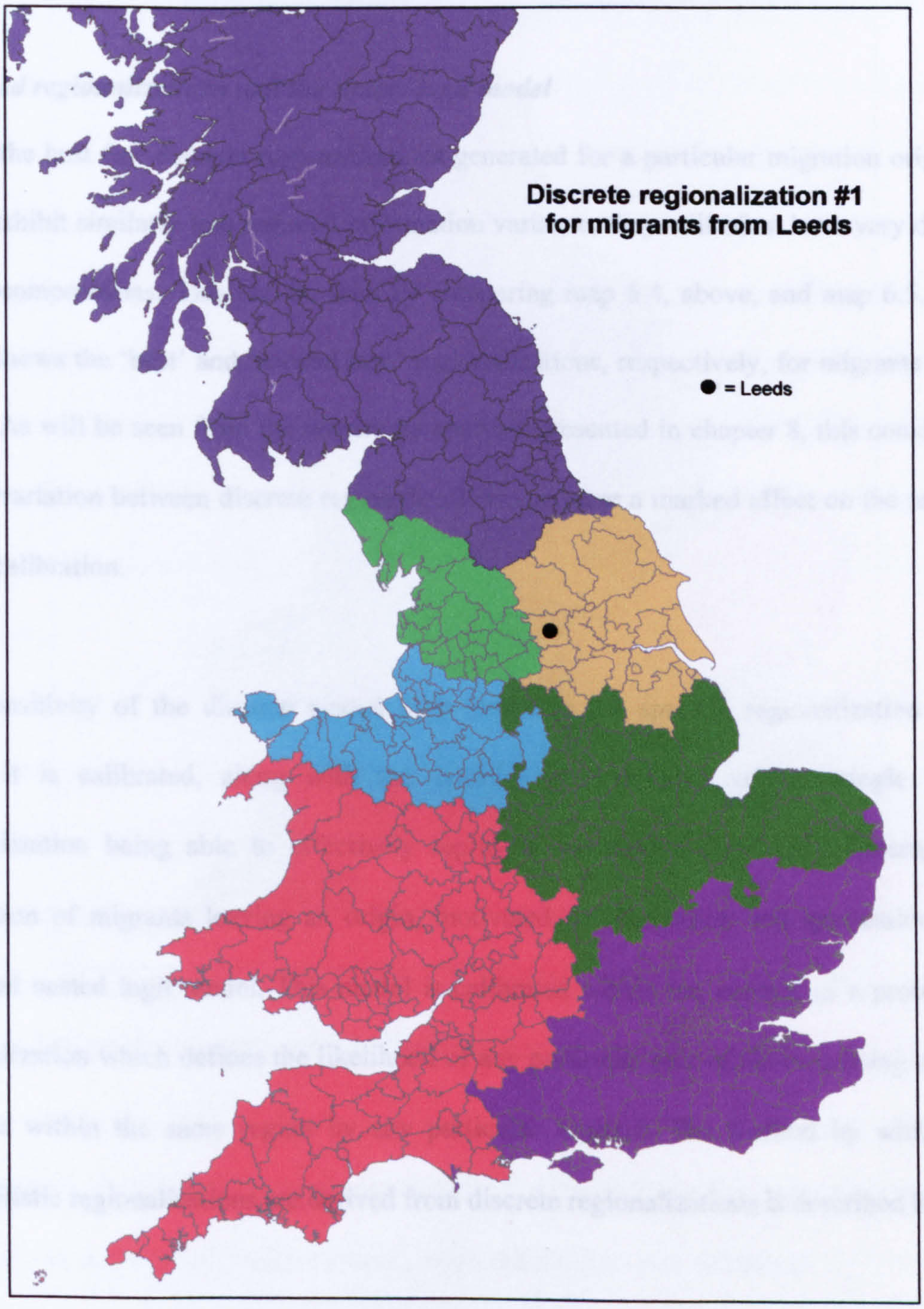
Step 10: Rank regionalizations by ascending regional information variance

In order to determine the best regionalizations the batch of 10,000 are now ranked according to the regional information variance values calculated in step 8, above.

Step 11: Select single 'best' regionalization for use in discrete nested logit models

After ranking the 10,000 valid regionalizations by regional information variance, the best single regionalization, i.e. the one with the lowest regional information variance, was selected as the regionalization to be used when calibrating the discrete nested logit model for migration from each particular origin. For the purposes of the weighted nested logit model the top 10% of these regionalizations are used, as will be discussed further below.

An example discrete regionalization is presented in map 6.4, below. This map shows the regionalization that had the lowest regional information variance of a batch of 10,000 regionalizations generated to represent the mental destination clustering of migrants leaving Leeds. It can be seen that, as predicted, there are smaller regions in the vicinity of Leeds, and larger regions further away in the far north and south of the country.



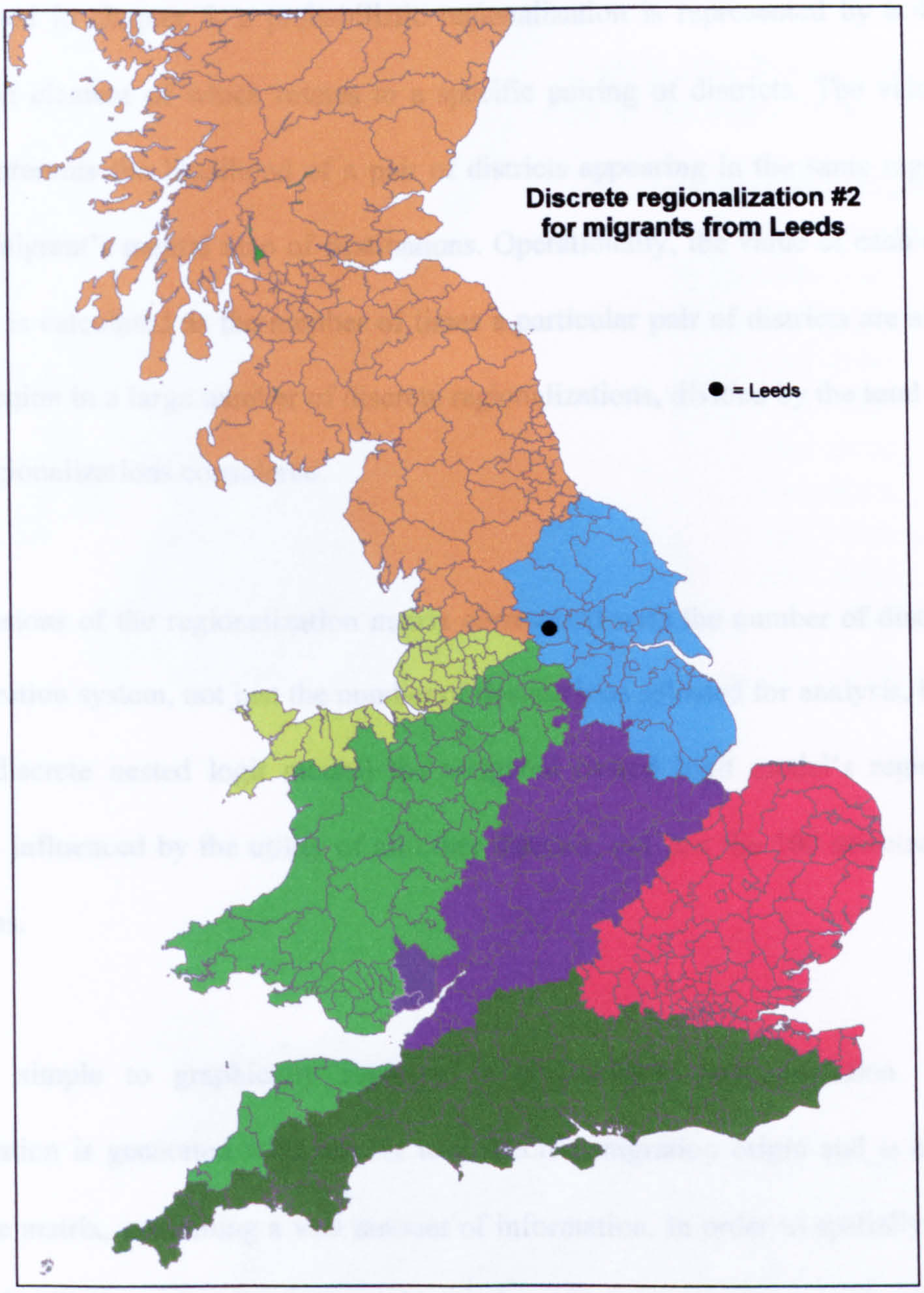
Map 6.4: 'Best' discrete regionalization for migrants leaving Leeds.

An extension to the nested logit framework, termed the weighted nested logit model, was introduced in the previous chapter, and was shown to utilize probabilistic regionalization definitions rather than discrete allocations of destinations to regions. The method by which such probabilistic regionalizations are generated through the aggregation of many discrete regionalizations is described below.

Weighted regionalizations and the nested logit model

Whilst the best few discrete regionalizations generated for a particular migration origin will likely exhibit similarly low regional information variance, they will often have very different spatial compositions. This can be seen by comparing map 6.4, above, and map 6.5, below, which shows the 'best' and 'second best' regionalizations, respectively, for migrants leaving Leeds. As will be seen from the sensitivity analysis presented in chapter 8, this considerable spatial variation between discrete regionalizations can have a marked effect on the results of model calibration.

This sensitivity of the discrete nested logit model to the specific regionalization against which it is calibrated, along with the intuitive implausibility of any single discrete regionalization being able to effectively represent the mental maps of a heterogeneous population of migrants leaving an origin, motivated the derivation and application of the weighted nested logit model. This model is calibrated within the context of a probabilistic regionalization which defines the likelihood of any particular pair of districts being cognized together within the same region by any particular migrant. The method by which these probabilistic regionalizations are derived from discrete regionalizations is described below.



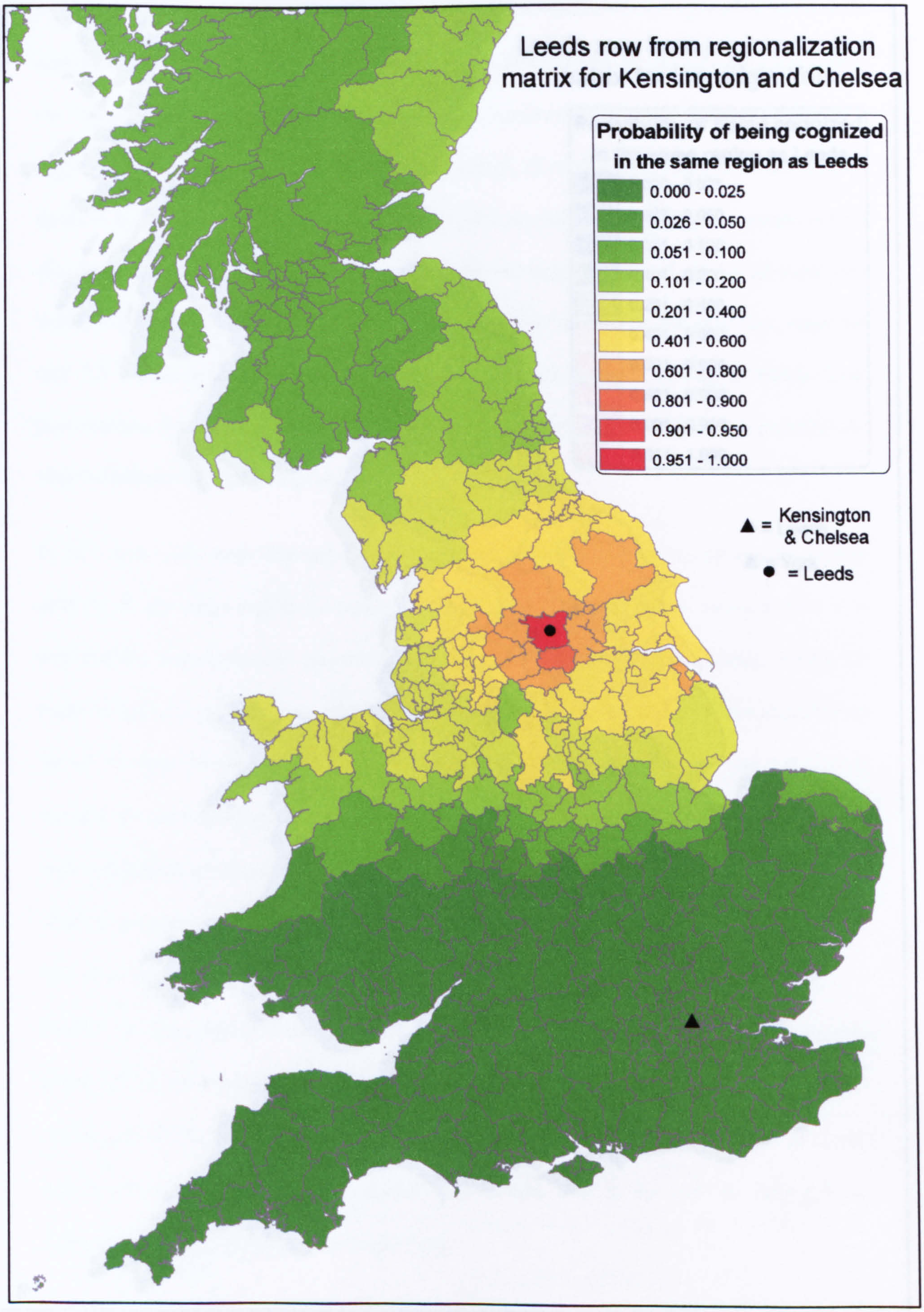
Map 6.5: Second 'best' discrete regionalization for migrants leaving Leeds.

In a discrete regionalisation each destination is allocated to one and only one region. However, in a weighted regionalisation, each destination has associated with it a set of probabilities that indicate the likelihood that a migrant will consider that destination in the same region as each other destination. The author proposes that this is likely to provide a closer approximation to the reality of migrants' cognitive hierarchies, as it accounts for the variety of mental maps that migrants with differing experiences, information levels and social networks must inevitably construct.

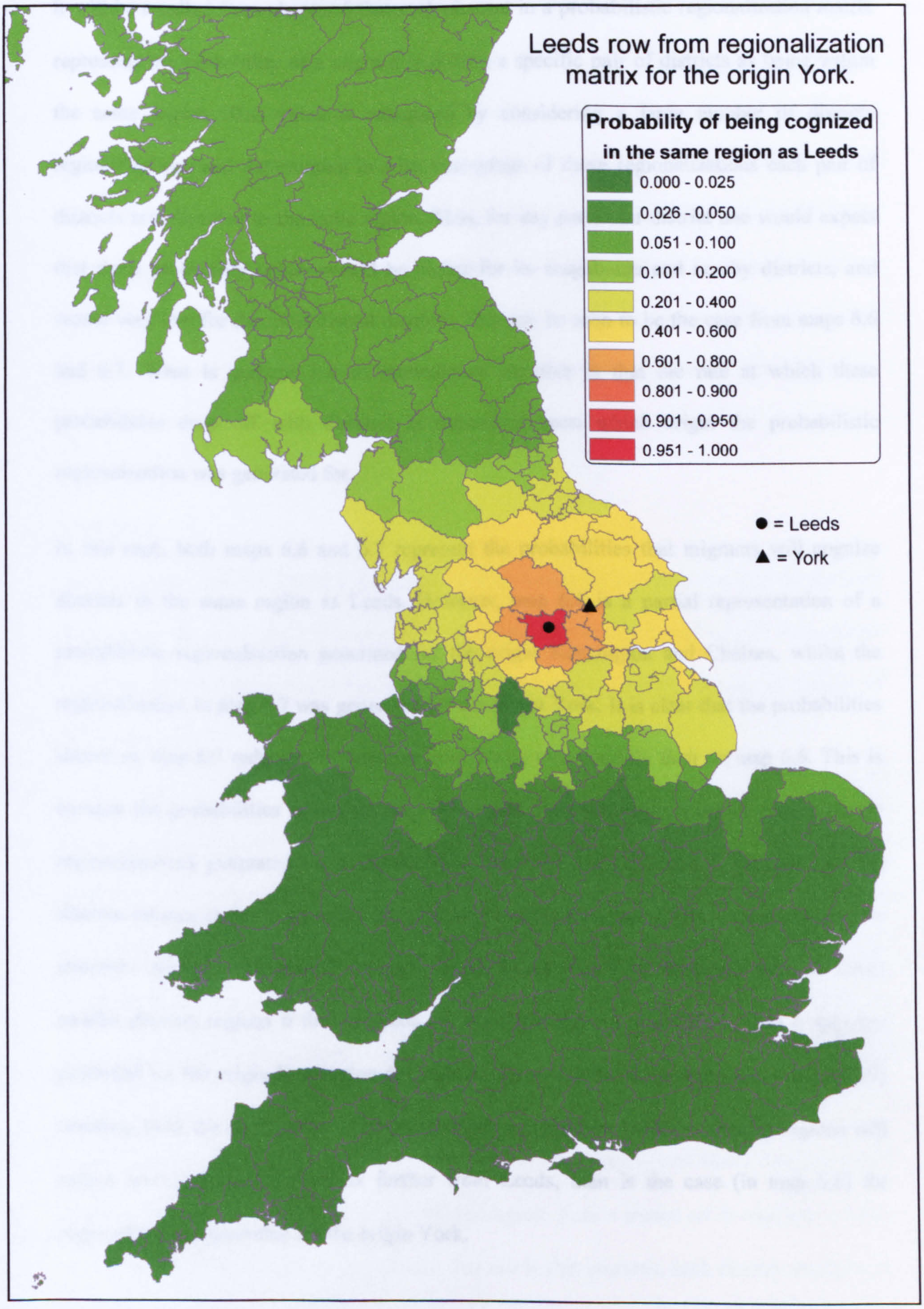
As mentioned in chapter 5, a probabilistic regionalization is represented by a 459-square matrix, each element of which relates to a specific pairing of districts. The value of each element represents the likelihood of a pair of districts appearing in the same region in any particular migrant's mental map of destinations. Operationally, the value of each element in this matrix is calculated as the number of times a particular pair of districts are allocated to the same region in a large number of discrete regionalizations, divided by the total number of discrete regionalizations considered.

The dimensions of the regionalization matrix correspond with the number of districts in the entire migration system, not just the number of destinations selected for analysis, because (as with the discrete nested logit model) the weighted nested logit model's regional utility variable is influenced by the utility of all other districts, not just the 100 selected migration destinations.

It is not simple to graphically represent a probabilistic regionalization. Each such regionalization is generated with respect to a specific migration origin and is essentially a 459-square matrix, containing a vast amount of information. In order to spatially represent a single probabilistic regionalization would require 459 separate maps, each one of which presents the information from a single row of the regionalization matrix. This is obviously not a useful way in which to visualize the regionalization as a whole, but it is useful to examine two such maps in order to understand how the probabilistic regionalization is constructed, and also how the weighted utility variable is calculated. Map 6.6, below, presents the row representing the district Leeds from a regionalization matrix created with respect to the origin Kensington and Chelsea. For comparative purposes, map 6.7, below provides a spatial view of the row representing the district Leeds, from a regionalization matrix created with respect to the origin York.



Map 6.6: Leeds row from regionalization matrix generated for origin Kensington & Chelsea.



Map 6.7: Leeds row from regionalization matrix generated for origin York.

It will be recalled from chapter 6 that each element in a probabilistic regionalization matrix represents the probability of a migrant cognizing a specific pair of districts as being within the same region. This value is calculated by considering a large number of discrete regionalizations and determining in what percentage of those regionalizations each pair of districts are allocated to the same region. Thus, for any particular district one would expect that these probability values would be higher for its neighbours and nearby districts, and would very low for the most distant districts. This can be seen to be the case from maps 6.6 and 6.7. What is perhaps not so immediately intuitive is that the rate at which these probabilities drop off with distance is dependent upon which origin the probabilistic regionalization was generated for.

In this case, both maps 6.6 and 6.7 represent the probabilities that migrants will cognize districts in the same region as Leeds. However, map 6.6 is a partial representation of a probabilistic regionalization generated for the origin Kensington and Chelsea, whilst the regionalization in map 6.7 was generated for the origin York. It is clear that the probabilities shown on map 6.7 reduce with distance from Leeds more rapidly than on map 6.6. This is because the probabilities shown on map 6.7 results from the aggregation of many discrete regionalizations generated for a district very close to Leeds. Chapter 6 explains that the discrete regions closer to an origin are generally smaller than those further away, in order to minimize regional information variance. Thus, Leeds will have generally been in much smaller discrete regions if those regions are generated for the origin York, than if they are generated for the origin Kensington & Chelsea. Consequently, the probabilities (in map 6.7) resulting from the aggregation of co-membership information for those smaller regions will reduce more rapidly for districts further from Leeds, than is the case (in map 6.6) for regionalizations generated for the origin York.

These probabilities are used when calculating the weighting that will be applied when the utility of each and every district is aggregated to produce the 'regional utility' value for Leeds. In spatial terms, this means the regional utility variable for Leeds resulting from a probabilistic regionalization generated for York (map 6.7) will represent a more concentrated region, than is the case for the regionalization generated for Kensington and Chelsea. The author argues that this is intuitively more appropriate as migrants from York are more likely to be able to differentiate between locations around Leeds, as they are so much closer, and their perception of Leeds will be less coloured by the characteristics of its surrounding areas, than that of migrants from more distance origins.

Summary

Appropriate and effective choice set definition is obviously of key importance in the analysis of any situation involving a discrete choice between alternatives. In the case of hierarchical migration destination choice defining the choice set means defining clusters or regions of destinations.

The research reported here provides evidence supporting the hierarchical theory of migration destination choice by demonstrating that modelling even a simple two-level hierarchy, or two step decision process, provides a better approximation to observed migration behaviour than do traditional 'flat processing' models. In such two-level analysis the choice set is simply a regionalisation of all the many individual destinations.

At their most simple these regionalizations are discrete regionalizations where each and every possible destination is contained in exactly one region. Such a choice set is required in order to calibrate the discrete nested logit model. For use in this research, such discrete regions are defined in such a way as to minimise the variance in 'total information' that a migrant from a specific destination is predicted to have about each region. So, for instance, regions more distant from a migrant's origin are likely to be larger as that migrant will *ceteris paribus* have less information about individual destinations that are further away.

A less rigid means of defining the choice set for such migration destination choice analysis is provided by the so called 'accessibility' statistic from competing destinations modelling. This is essentially a measure of the likelihood that a specific destination will be grouped with other potential destinations in any particular migrant's mental map. This has the advantage of not imposing one specific regionalization on all migrants when in reality every migrant is likely to mentally group destinations into a different hierarchy when deciding on a migration destination.

The weighted nested logit model attempts to deliver aspects of both of these approaches by defining a probabilistic regionalization based on the aggregation of a large number of discrete regionalizations to form a matrix representing the likelihood of each possible pairing of destinations appearing in the same region in any particular migrant's mental map of destinations. This probabilistic regionalization definition is used by the weighted nested logit model to calculate a probabilistically weighted regional utility variable that is unique for every potential migration destination (unlike the discrete nested logit model's regional utility variable, which has the same value for all destinations allocated to the same discrete region).

This probabilistic regionalization is also employed in the calibration of the hybrid weighted nested logit model, which combines the benefits of the competing destinations and weighted nested logit models by including the hierarchical variables from both models: the accessibility and weighted nested logit variables.

The chapters in this section have described the various data, hierarchical models and choice sets that are applied here to the analysis of migration destination choice. The chapters in the following section report results obtained from the calibration of these hierarchical migration models and compares them with results from the non-hierarchical traditional migration model. Chapter 7 compares the results obtained from calibrating the competing destinations model with those from the tradition migration destination choice model; chapter presents similar comparisons for the nested logit models; chapter 9 reports on more direct comparisons between the various hierarchical models introduced here; and, chapter 10 uses results from all the various models to examine age, gender and marital status variation in migration destination choice behaviour.

Chapter Seven

Competing Destinations Model

This chapter presents the results obtained from the calibration of the competing destinations migration destination choice model and compares and contrasts these results with those obtained from the calibration of a traditional 'flat-processing' logit model (defined in equation 2.4). The comparison focuses on the goodness-of-fit of the models' predicted values to observed data, and on the similarities and differences between the parameter estimates resulting from the calibration of the two models. The model results presented in this chapter are based on calibrations against the observed behaviour of all migrants aged 16 and over. Comparisons of the migration destination choice behaviour of migrants disaggregated by age, gender and marital status are presented in chapter 10.

The competing destinations model, like the traditional model, does not have any dependency upon an origin-specific regionalization, or any other variable that is derived in an origin-specific manner. This enables the model to be calibrated globally as well as independently for each origin. However, it will become apparent in the final section of this chapter that the results from a single global model calibration that attempts to simultaneously predict the migration behaviour of ALL migrants conceals a great deal of spatial variation in the behaviour of migrants from different origins. Because of this, the next two sections of this chapter, that deal with goodness-of-fit and parameter estimate variation, respectively, will make use of the results from origin-specific model calibrations of both the competing destinations and traditional models. Global and local origin-specific results are compared and discussed further at the end of this chapter.

It is useful to recall that in practical terms the competing destinations model can be considered to be a traditional 'flat-processing' model with an additional explanatory variable, the accessibility statistic. Though their derivation and theoretical underpinnings differ, operationally they are similar models. Thus, on one level, this chapter can be considered to be an examination of the effects of adding an explanatory variable to the traditional model of migration destination choice. One consequence of this is that no calibration of a competing destinations model should exhibit a worse goodness-of-fit, as indicated by its R^2, than a traditional model calibrated for migration from the same origin against comparable observed data. Even in a 'worst case' scenario, where the additional accessibility variable has no correlation whatsoever with observed migration destination choice behaviour, the accessibility variable's parameter estimate will turn out to be statistically no different from zero, meaning that the new variable can essentially be dropped from the model as it has no effect on the goodness-of-fit of the model as a whole.

Goodness of model fit

This section examines the predictive ability of the competing destinations and traditional models of migration destination choice. It is interesting to compare goodness-of-fit at a number of levels: between models; between migrant origins; and, between migrant destinations. Comparison between models addresses the central research question of this thesis: whether migration destination choice is a hierarchical process. Spatial variation becomes apparent when goodness-of-fit measures from a set of origin-specific model calibrations are shown on a map, supporting the argument for local calibration and interpretation of migration models. For a specific origin-specific model calibration examination of how the predictive ability of the model varies between individual destinations can also highlight interesting spatial variation in migration behaviour which can potentially inform model specification.

In this research the goodness-of-fit of the models has been evaluated in three ways, examining: R^2 statistics, Akaike Information Criterion (AIC) statistics and flow residuals. R^2 and AIC statistics are useful for comparisons between models and between migrant origins. Flow residuals offer a way to examine the spatial patterns in the accuracy of predicted flows to various migration destinations from a particular origin, and also to investigate how these spatial patterns of goodness-of-fit vary between origin-specific model calibrations.

Coefficient of Determination - R^2 and R^2_{adj}

The Coefficient of determination, or R^2, represents the proportion of the variance in the observed data that is explained by a model's explanatory variables. The R^2 statistic ranges in value from 0 to 1, with a higher value indicating that a model's predicted migration flows more closely correlate with observed flows. The definition of the R^2 statistic is shown below in equation 7.1:

$$R^2 = 1 - \frac{RSS}{TSS} \qquad \text{(Eq.7.1a)}$$

Where:

RSS is the sum of the squares of the residuals:

$$ESS = \sum_{i=1}^{n} (m_i - \hat{m}_i)^2 \qquad \text{(Eq.7.1b)}$$

TSS is the sum of the squares of the observed variation:

$$TSS = \sum_{i=1}^{n} (m_i - \bar{m})^2 \qquad \text{(Eq.7.1c)}$$

m_i = observed migration to destination i

$\hat{m}_i$ = predicted migration to destination i

$\bar{m}$ = mean observed migration

n = number of observations

Equation 7.1: Definition of the coefficient of determination, R^2.

It is a characteristic of the R^2 statistic that its value for a particular model cannot be reduced through the addition of a new explanatory variable. When comparing models the inclusion of explanatory variables that add nothing to the model's predictive ability should be considered a detrimental characteristic and it is therefore useful to employ a goodness-of-fit measure that takes into account the complexity of the model. Having such superfluous explanatory variables in a model's specification increases the chances of introducing multicollinearity which can bias parameter estimates and greatly complicate their interpretation. The adjusted R^2 takes into account model complexity by including the number of explanatory variables in its definition:

$$R^2_{adj} = 1 - \frac{RSS/(n-p-1)}{TSS/(n-1)} \qquad (Eq.7.2)$$

Where:

RSS, TSS and n are defined as for equation 7.1 above, and,
P = the number of explanatory variables in the model

Equation 7.2: Definition of the adjusted R^2 statistic, (R^2_{adj}).

It can be seen from equations 7.1 and 7.2 that for any particular calibration of a model R^2_{adj} will be lower than R^2. The difference will be greater for more complex models with more explanatory variables, and, as can be seen in figure 7.1 below, is greater for badly fitting models. Figure 7.1 below compares the R^2_{adj} and R^2 for a set of 100 origin-specific calibrations of a traditional migration destination choice model.

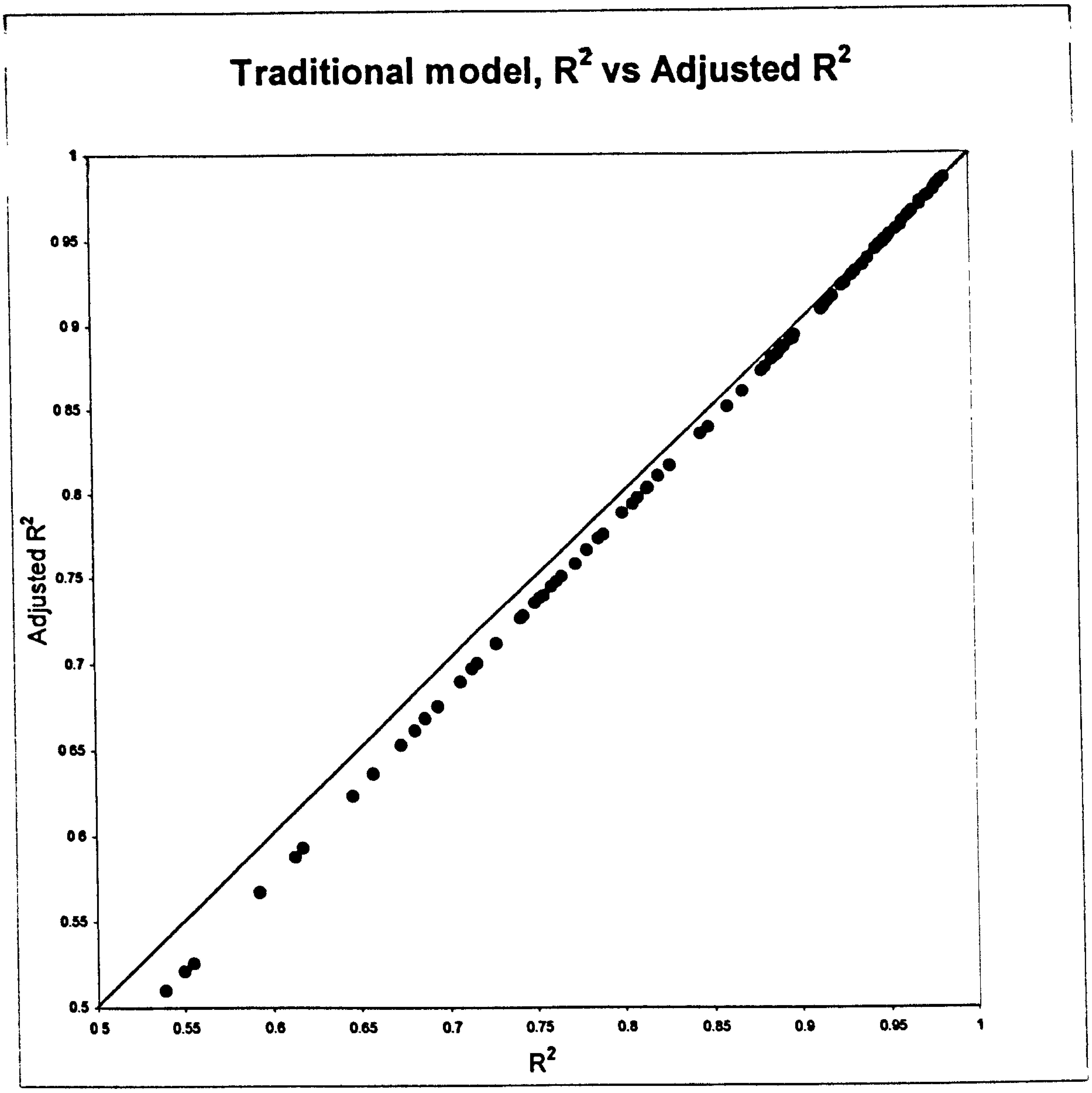


Figure 7.1: R^2_{adj} and R^2 for the traditional model.

Because of the beneficial characteristics described above and also the greater acceptance of the adjusted R^2 statistics for comparisons between differently specified models, it is this statistic which is presented in all discussions of R^2 throughout the remainder of this thesis.

Figure 7.2 below plots R^2_{adj} values to compare the goodness-of-fit of the traditional and competing destinations models.

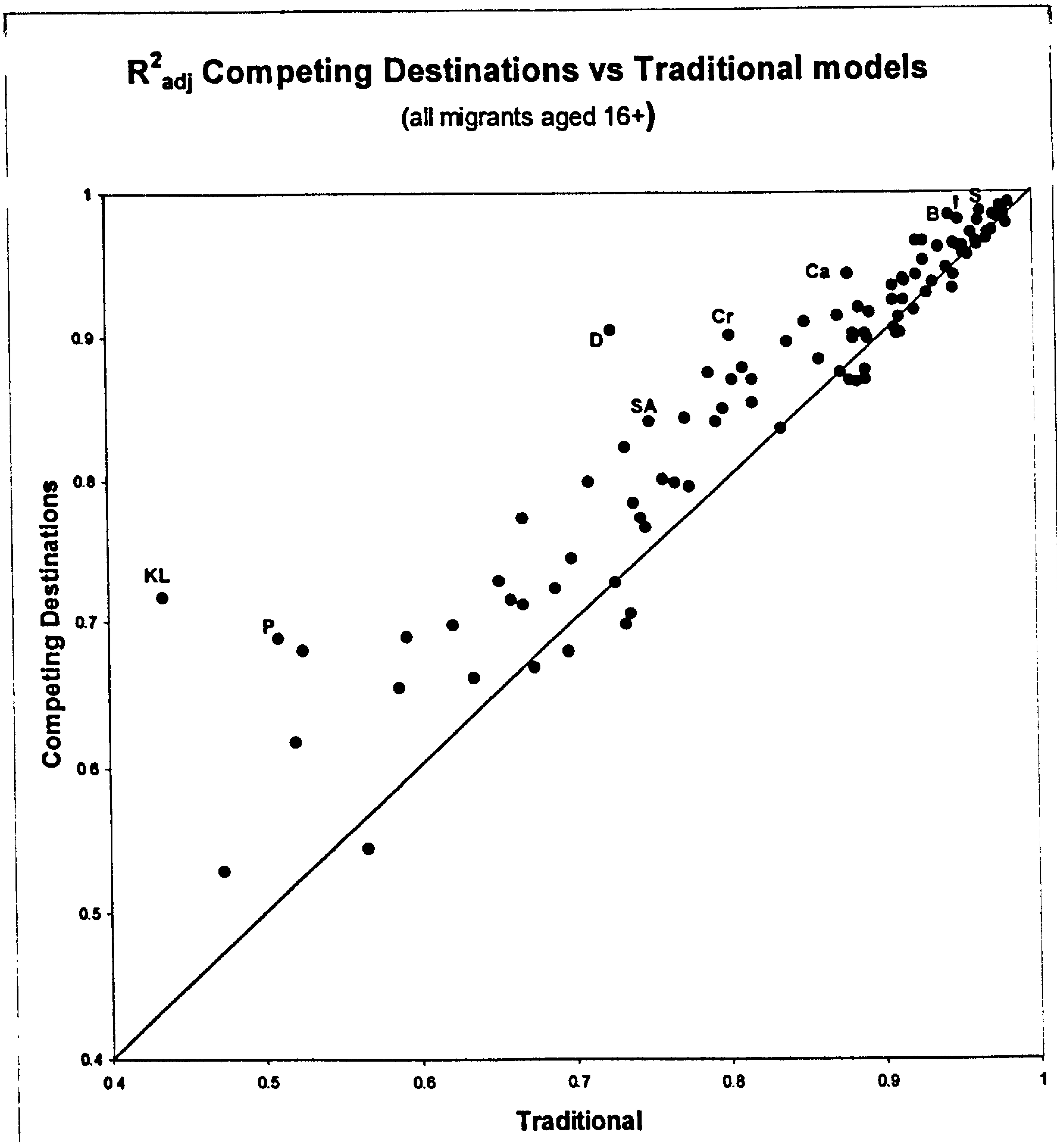


Figure 7.2: R^2_{adj}, traditional and competing destinations models, (outliers labelled).

It can be seen from figure 7.2 that calibrations of the competing destinations model generally result in higher R^2_{adj} values than for the traditional model. This indicates that in most cases the competing destinations model predicts more of the variation in the observed migration flow data than the traditional model. Given that in operational terms the only difference between these two models is the competing destinations model's additional accessibility variable, it is evident that this accessibility variable is explaining some of the variation in the observed migration flow data that remains unexplained by the traditional 'flat-processing' model.

Akaike Information Criterion (AIC)

Another commonly used statistic when comparing model performance is the Akaike Informatino Criterion (AIC). Its definition is based on the sum of squared error residuals, but like the R^2_{adj}, it also takes into account the complexity of the model under consideration. The definition of the AIC statistic is presented in equation 7.3 below.

$$AIC = 2k + n\ln(RSS/n) \qquad \text{(Eq. 7.3)}$$

Where:

k is the number of parameters in the model,

n is the number of observations,

RSS is the sum of the squares of the residuals, see eq.7.1b above.

Equation 7.3: Akaike Information Criterion, (AIC).

Note that, unlike the R^2_{adj}, whose definition includes the squared variation in the observed data in the denominator in order to produce a value bounded between 0 and 1, the AIC has no such denominator and does not have an upper bound. The results of this is that absolute values of AIC are meaningless when comparing models calibrated against different observed data – as it is impossible to determine whether any difference in the values results from better or worse model fit, or whether it results from differing variation in the observed data against which the models are calibrated. This is confirmed by the lack of any discernable relationship in figure 7.4 which plots R^2_{adj} values against AIC values for 100 origin-specific tradition model calibrations.

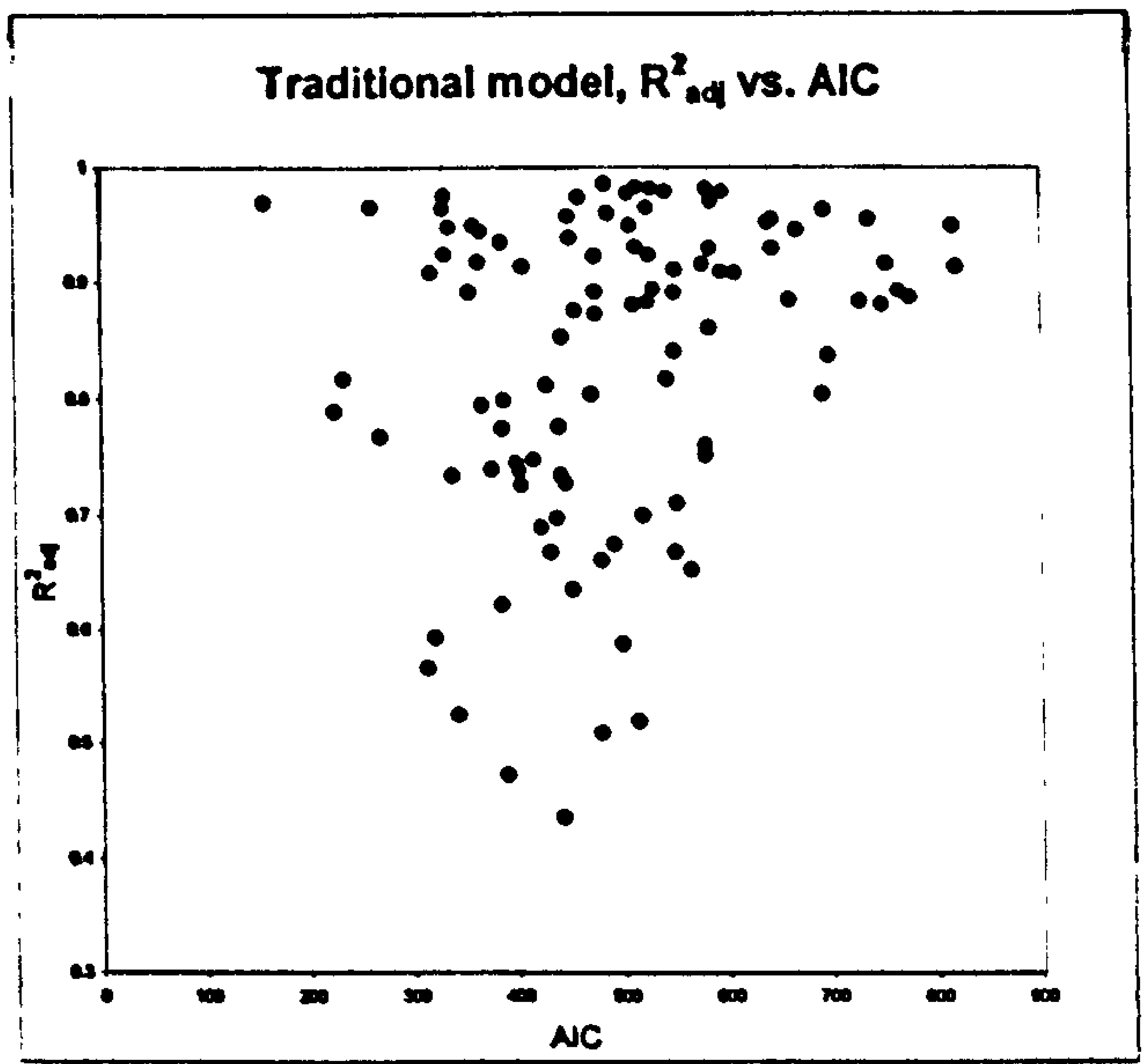


Figure 7.4: AIC vs R^2_{adj} for 100 origin-specific tradition model calibrations

The AIC statistic is intended for use when comparing two or more models that are calibrated against the same observed data. When used in this way improved goodness-of-fit is reflected by lower AIC values. So whilst it is not meaningful to compare absolute AIC and R^2_{adj} values for a set of different models, it is meaningful to compare the change in AIC and R^2_{adj} values between models when the models are calibrated against same observed migration data. Figure 7.5 compares the absolute change in AIC and R^2_{adj} values between origin-specific calibrations of the traditional and competing destinations models.

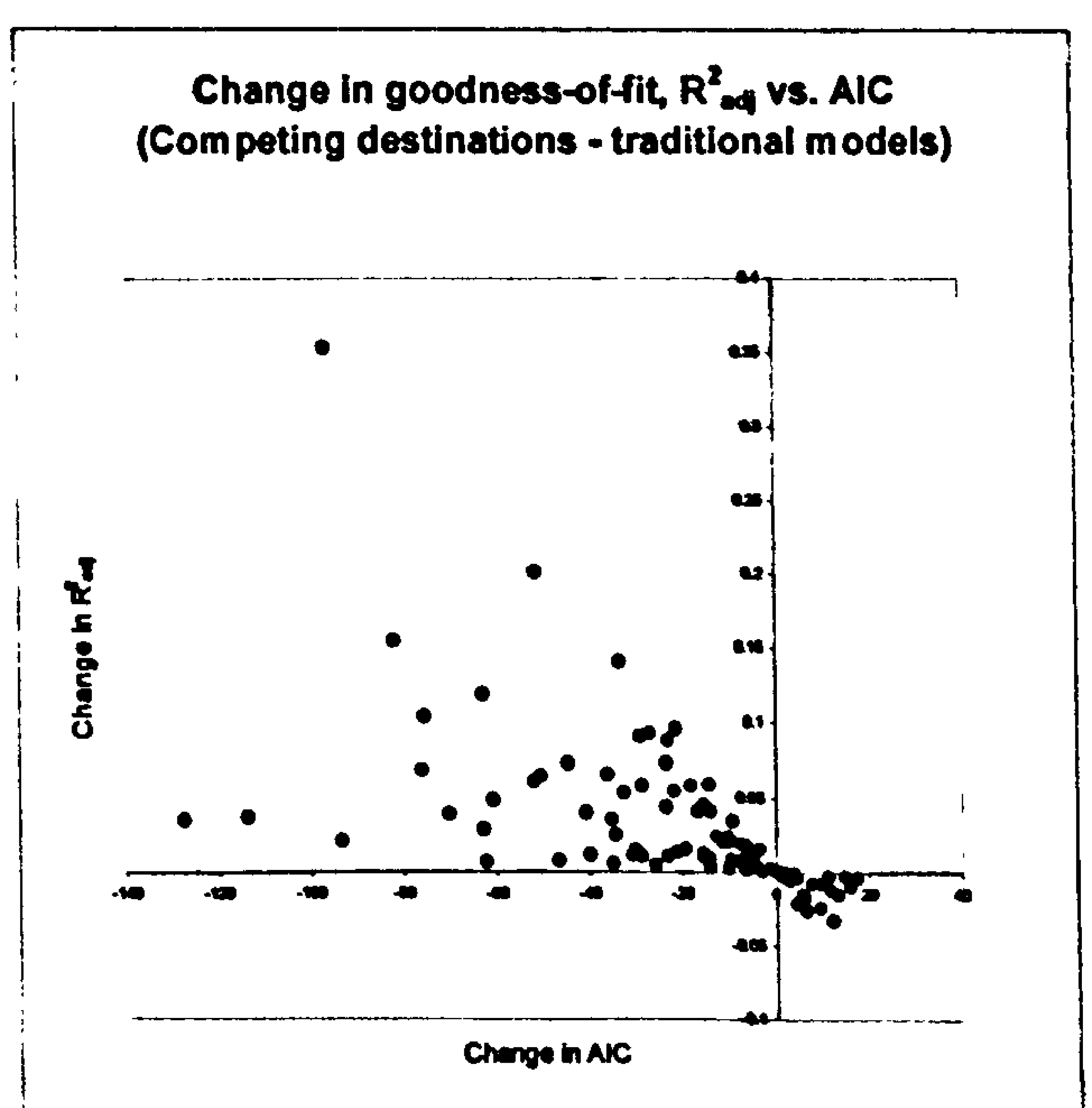


Figure 7.5: Change in AIC and R^2_{adj}, competing destinations - traditional models.

As one would expect, given the definitions of the two goodness-of-fit statistics, figure 7.5 shows a negative relationship between the changes in their values. Without exception, a reduction in AIC is associated with an increase in R^2_{adj}, and increases in AIC are associated with reductions in R^2_{adj} values.

Figure 7.6 below presented the changes in the AIC statistics for 100 origin-specific calibrations of the traditional and competing destinations models.

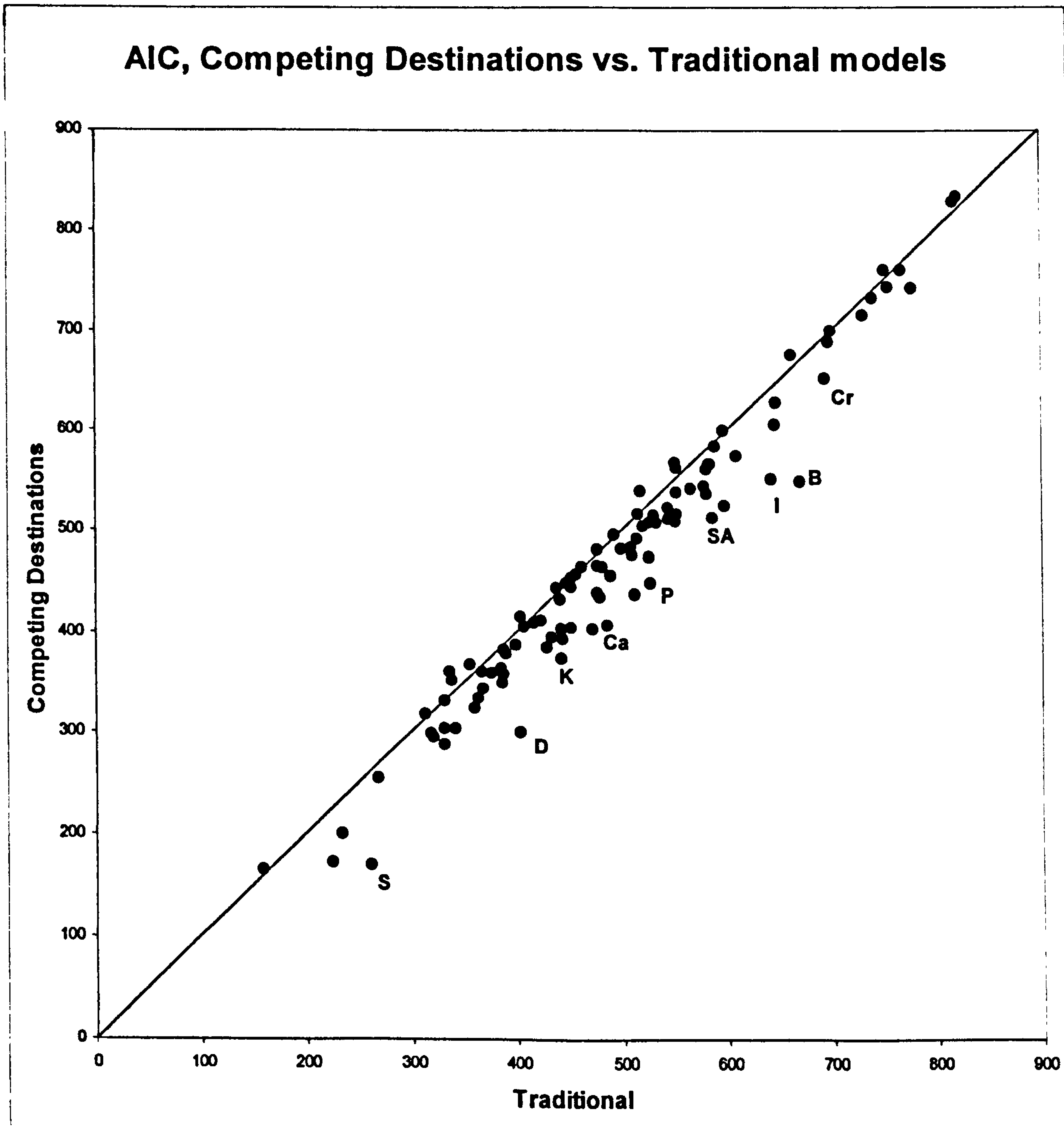


Figure 7.6: AIC: competing destinations vs. tradition models, (outliers labelled).

Recalling that better goodness-of-fit is reflected by a lower value of the AIC statistic – it is clear that figure 7.6 confirms the finding from comparison of R^2_{adj} values (in figure 7.2 above) that the competing destinations model provides improved goodness-of-fit over the tradition model.

The statistical distribution of the goodness-of-fit change between the traditional and competing destinations models is plotted in figure 7.7 below. This clearly shows that the model makes improvements in goodness-of-fit (i.e. reductions in AIC values) for the vast majority of migrant origins.

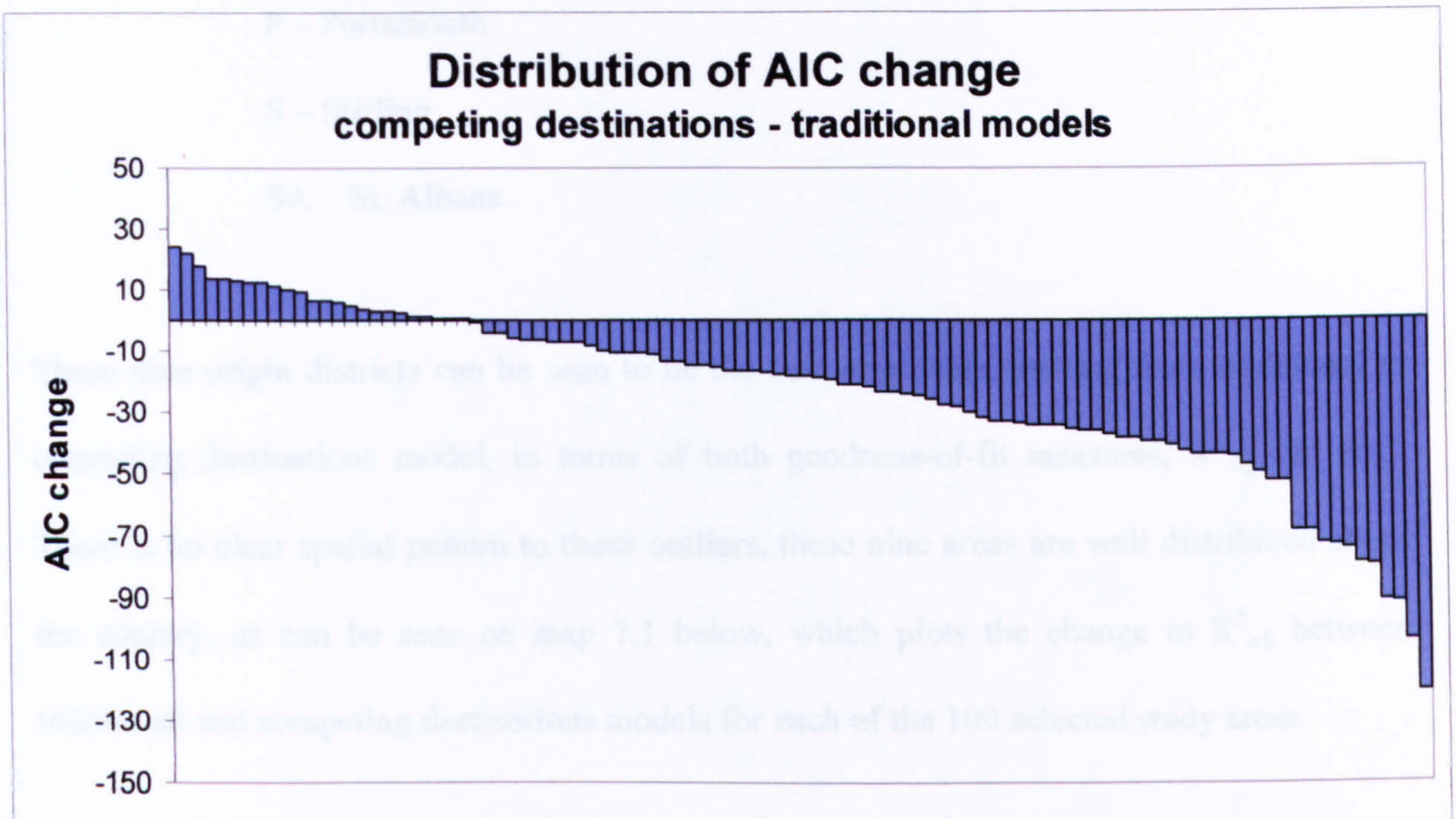


Figure 7.7: Distribution of change in AIC, competing destinations – traditional models.

It should be noted that for some migrant origins the AIC values are actually higher for the competing destinations model than for the traditional model. In these instances any small improvements in absolute model fit resulting from the introduction of the accessibility variable are more than offset by the increase in model complexity.

Outliers from examination of AIC and R^2_{adj}

Figures 7.2 and 7.6, comparing the R^2_{adj} and AIC values, respectively, resulting from calibrations of the traditional and competing destinations models, include character symbols beside a number of outlier data points in order that they can be identified as:

B – Bradford

Ca – Canterbury

Cr – Croydon

D – Derby

I – Islington

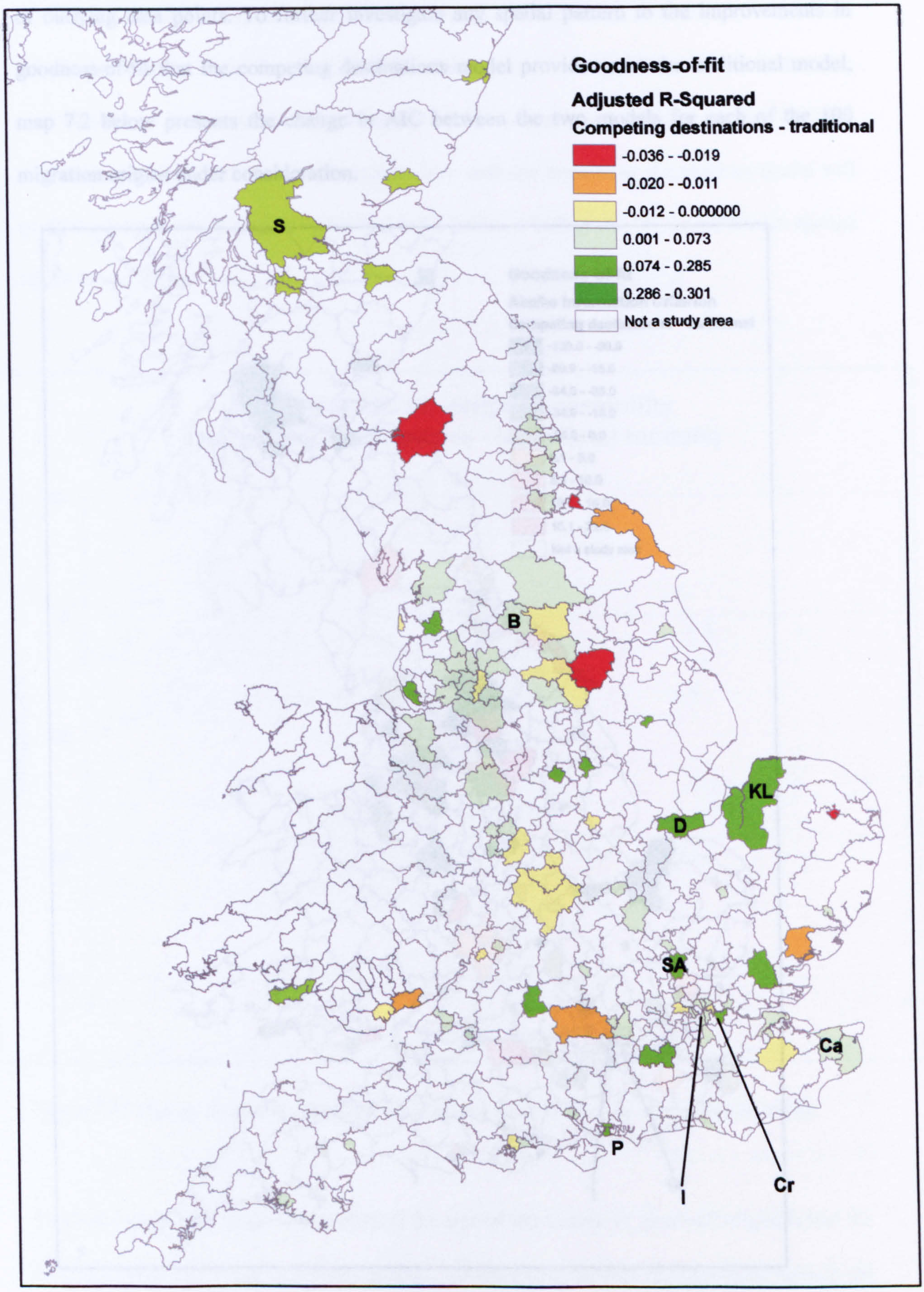
KL – Kings Lynn and West Norfolk

P – Portsmouth

S – Stirling

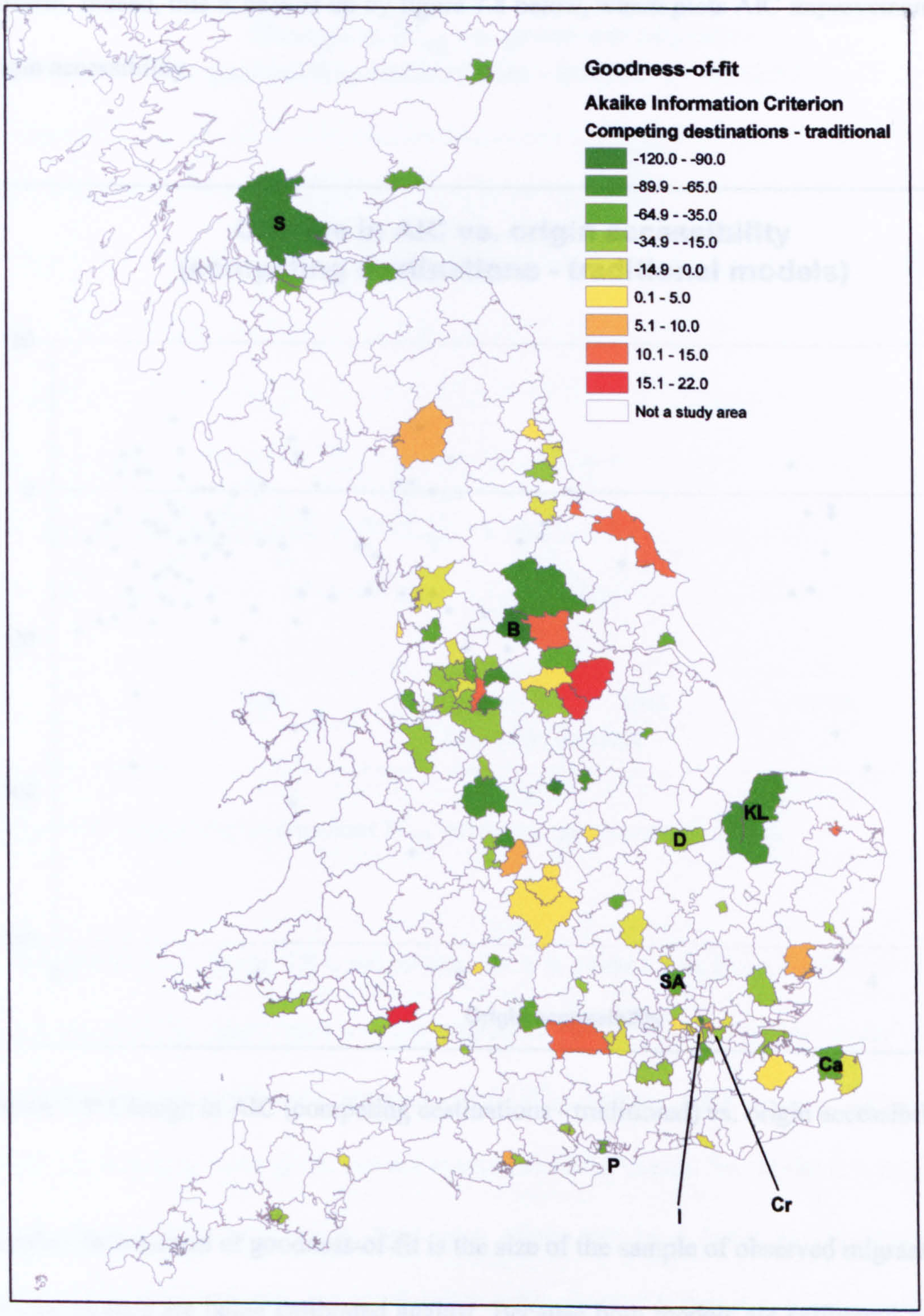
SA – St. Albans

These nine origin districts can be seen to be the best improvers, moving from traditional to competing destinations model, in terms of both goodness-of-fit measures, R^2_{adj} and AIC. There is no clear spatial pattern to these outliers, these nine areas are well distributed about the country, as can be seen on map 7.1 below, which plots the change in R^2_{adj} between traditional and competing destinations models for each of the 100 selected study areas.



Map 7.1: R^2_{adj} change, competing destinations - traditional models

It is not reasonable to draw any general conclusions from examination of this small number of outlying data points. To further investigate any spatial pattern to the improvements in goodness-of-fit that the competing destinations model provides over the traditional model, map 7.2 below presents the change in AIC between the two models for each of the 100 migration origins under consideration.



Map 7.2: AIC change, competing destinations – traditional models.

The size of the difference in goodness-of-fit between competing destinations and traditional models is indicated using colour coding. A slight trend is apparent towards more significant R^2_{adj} improvements in the South East of the country, but these best improvers include some coastal areas as well as districts in the home counties, suggesting that it is not the centrality of the migrant origins alone that determines how well the competing destinations model will perform. Indeed, this is backed up by figure 7.8 below, which plots AIC improvement against origin accessibility.

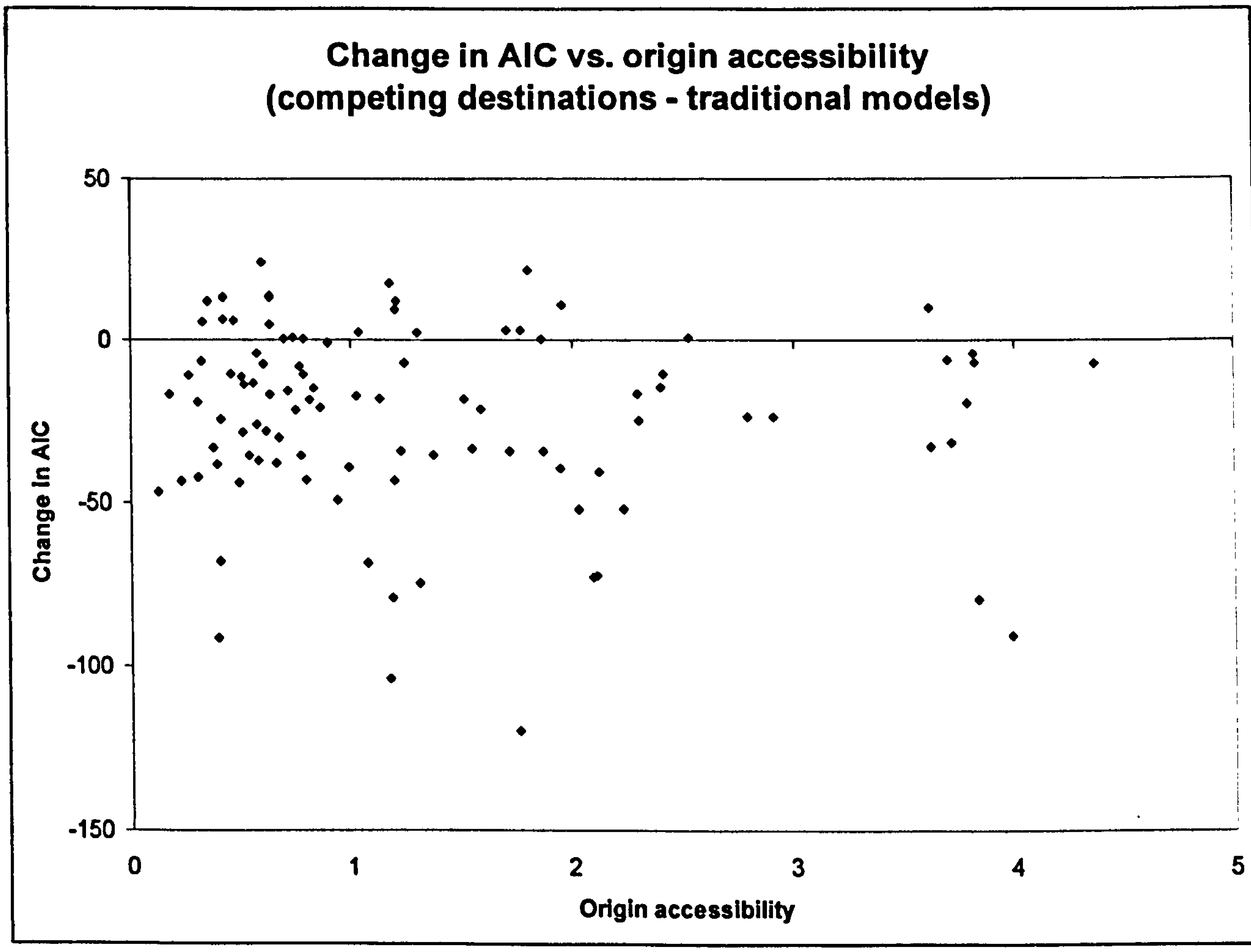


Figure 7.8: Change in AIC (competing destinations - traditional) vs. origin accessibility.

Another determinant of goodness-of-fit is the size of the sample of observed migrants that the various models are being calibrated against. Because both models are calibrated against the same empirical data this has no impact upon comparisons between models. Nonetheless, it is interesting to examine the relationship between the level of gross out migration of each origin and the R^2_{adj} values that result from origin-specific calibrations of the migration models. Figure 7.9 below plots R^2_{adj} values from origin-specific calibrations of the competing destinations model against the gross out migration from each district.

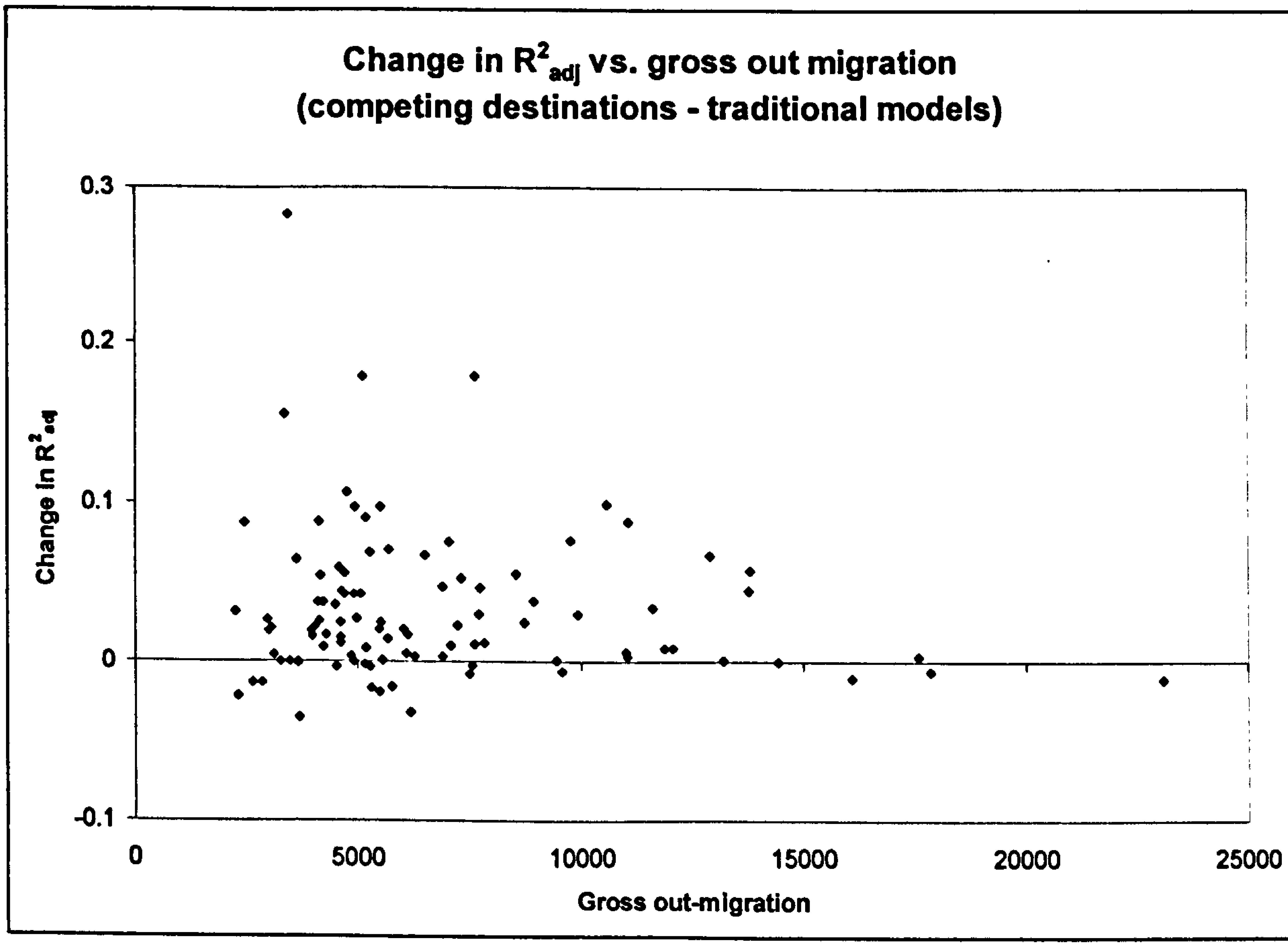


Figure 7.9: Competing destinations R^2_{adj} values vs. gross out-migration.

The relationship in figure 7.9 is not strong, but it is evident that those districts with the worst goodness-of-fit do indeed have low out-migration. However, there are also many origins with low out-migration that have extremely high R^2_{adj} values. Also, origins with higher out-migration appear to have good, but not exceptional R^2_{adj} values. So, whilst it is inevitable that the level of observed migration from each origin will be a factor in the accuracy of the results, it is clear from the very weak relationship figure 7.9 that it is not the limiting factor, and in no way invalidates the results of this research.

Flow Residuals

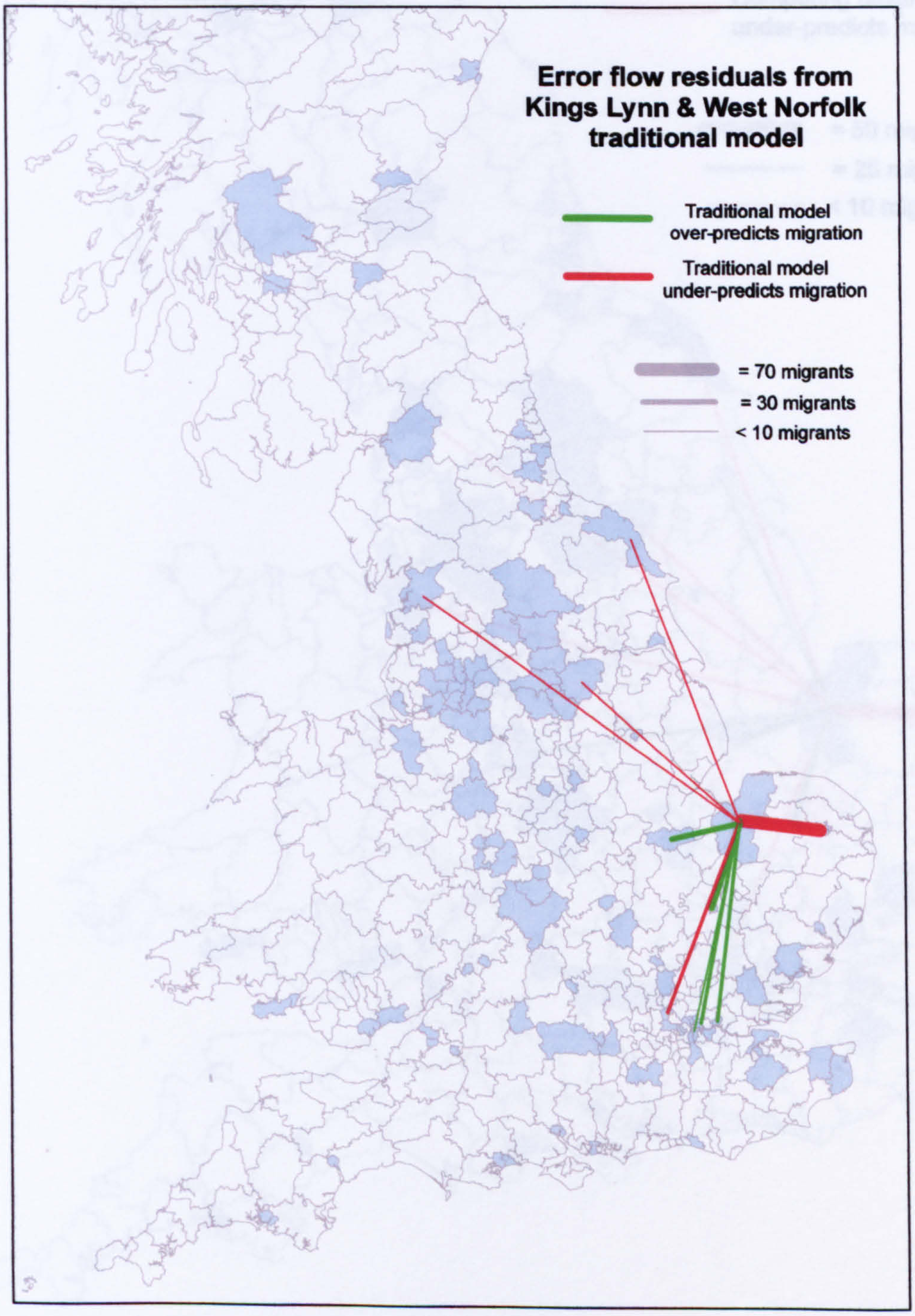
In addition to spatial variation in the goodness-of-fit of model calibrations from the various migration origins, there is also spatial variation in the accuracy with which a particular origin-specific model calibration predicts migration to the various potential migration destinations. Such spatial variation in a model's predictive ability across destinations is best visualized using flow residuals maps.

A flow residual is the difference between two flows between the same origin and destination. We are concerned here with the goodness-of-fit of the models, so we map the error flow residuals between a model's predicted flows and the observed migration flows against which that model was calibrated. Maps 7.3 And 7.4 below shows such error flow residuals between observed migration from Kings Lynn & West Norfolk and the migration flows predicted from that origin by the traditional and competing destinations models, respectively. In order to reduce visual congestion only residual flows larger than one standard deviation are plotted. Line thickness on all residual flow maps is proportional to the size of the residual flow.

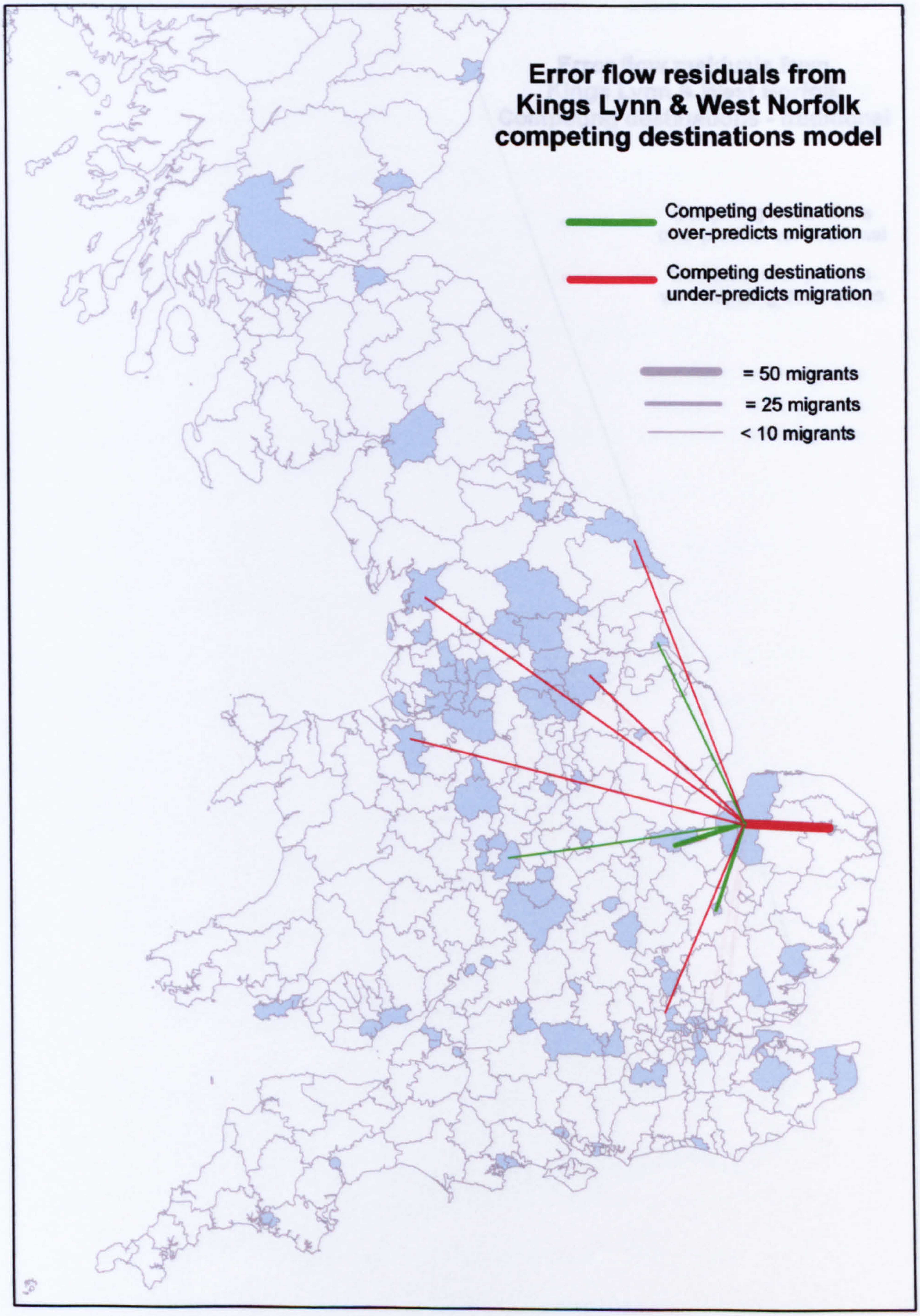
It is immediately clear from maps 7.3 and 7.4 that the general pattern of predicted migration from Kings Lynn and West Norfolk is similar for both the competing destinations and traditional models. For this reason, the residual between the predicted flows from the two models are compared directly on map 7.5 below. As in maps 7.3 and 7.4, only residuals greater than one standard deviation are plotted, and line thickness is proportional to flow residual size.

The key spatial differences between the migration predictions of the two models are immediately evident from map 7.5. The two models differ significantly in their predictions of migration from Kings Lynn and West Norfolk to a large cluster of quite local destinations almost due East of Kings Lynn and West Norfolk.

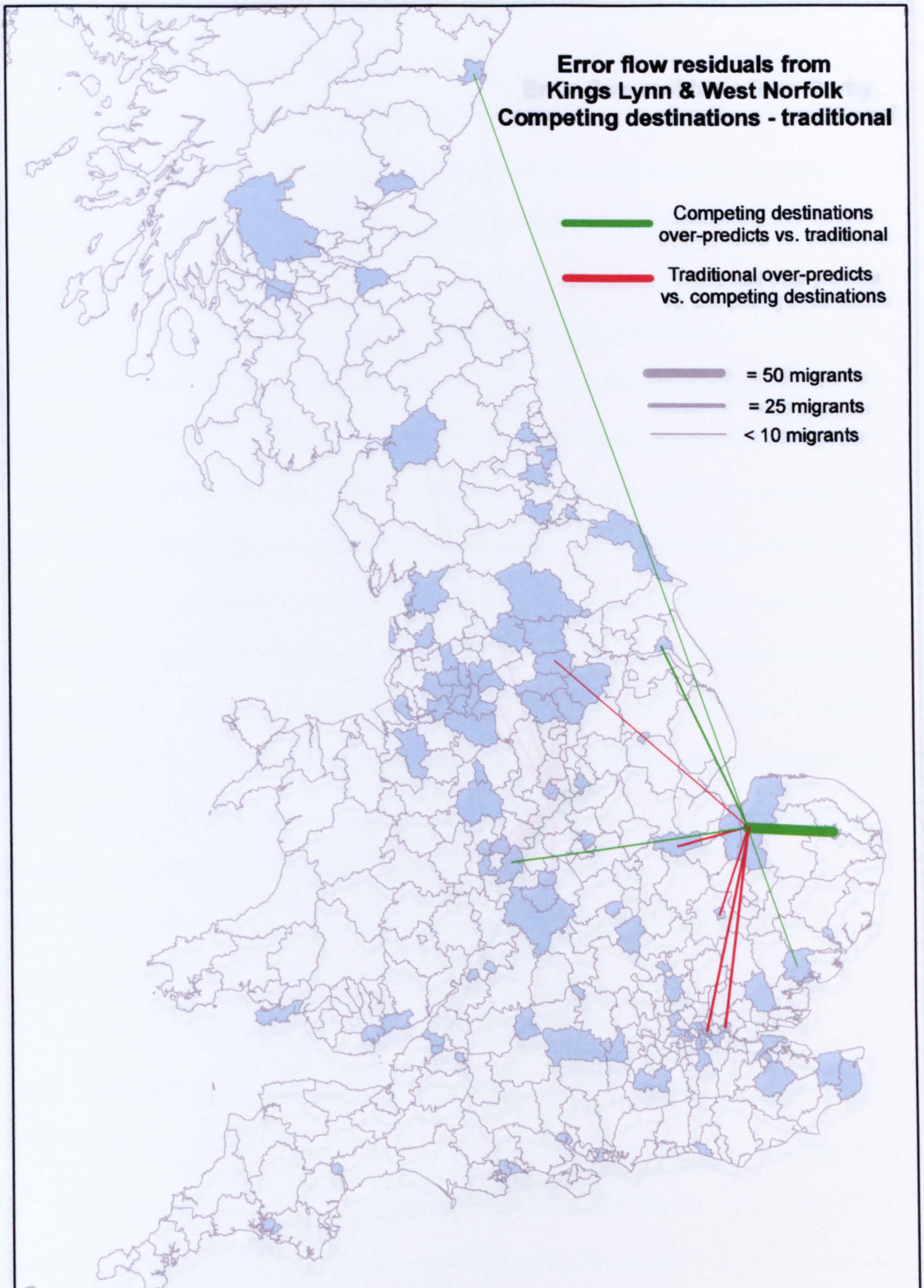
Such flow residuals can be mapped for each of the 100 selected migrant origins. However, it is most interesting to examine the residuals of the migration flows predicted by the models from those origins for which the addition of the accessibility variable results in the most dramatic improvement in goodness-of-fit. Here, two additional outliers from figures 7.2 and 7.6 are considered, and error flow residual maps for Derby and Portsmouth are presented below in maps 7.6 and 7.7, respectively.



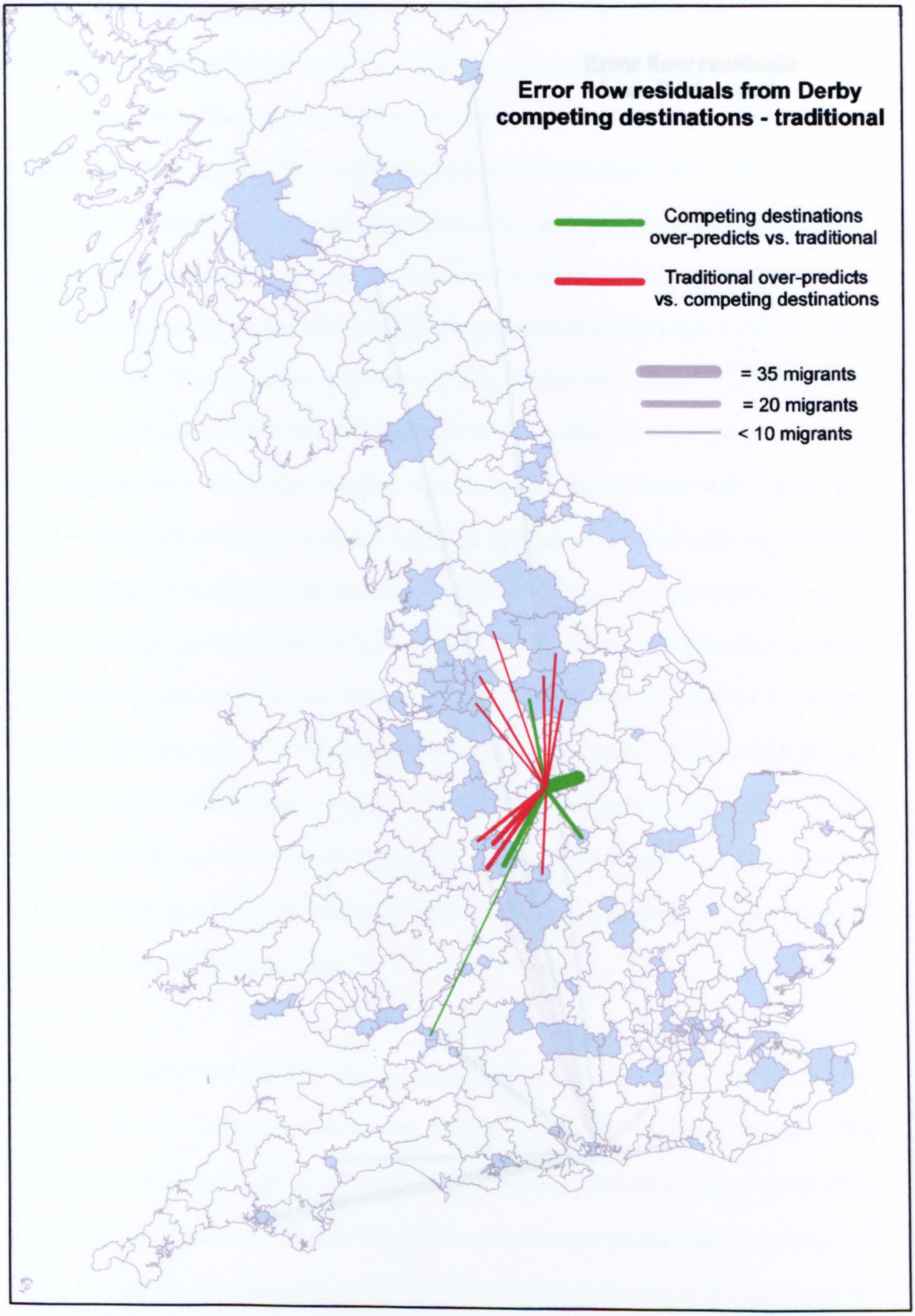
Map 7.3: Error flows residuals from Kings Lynn & W. Norfolk, traditional model.



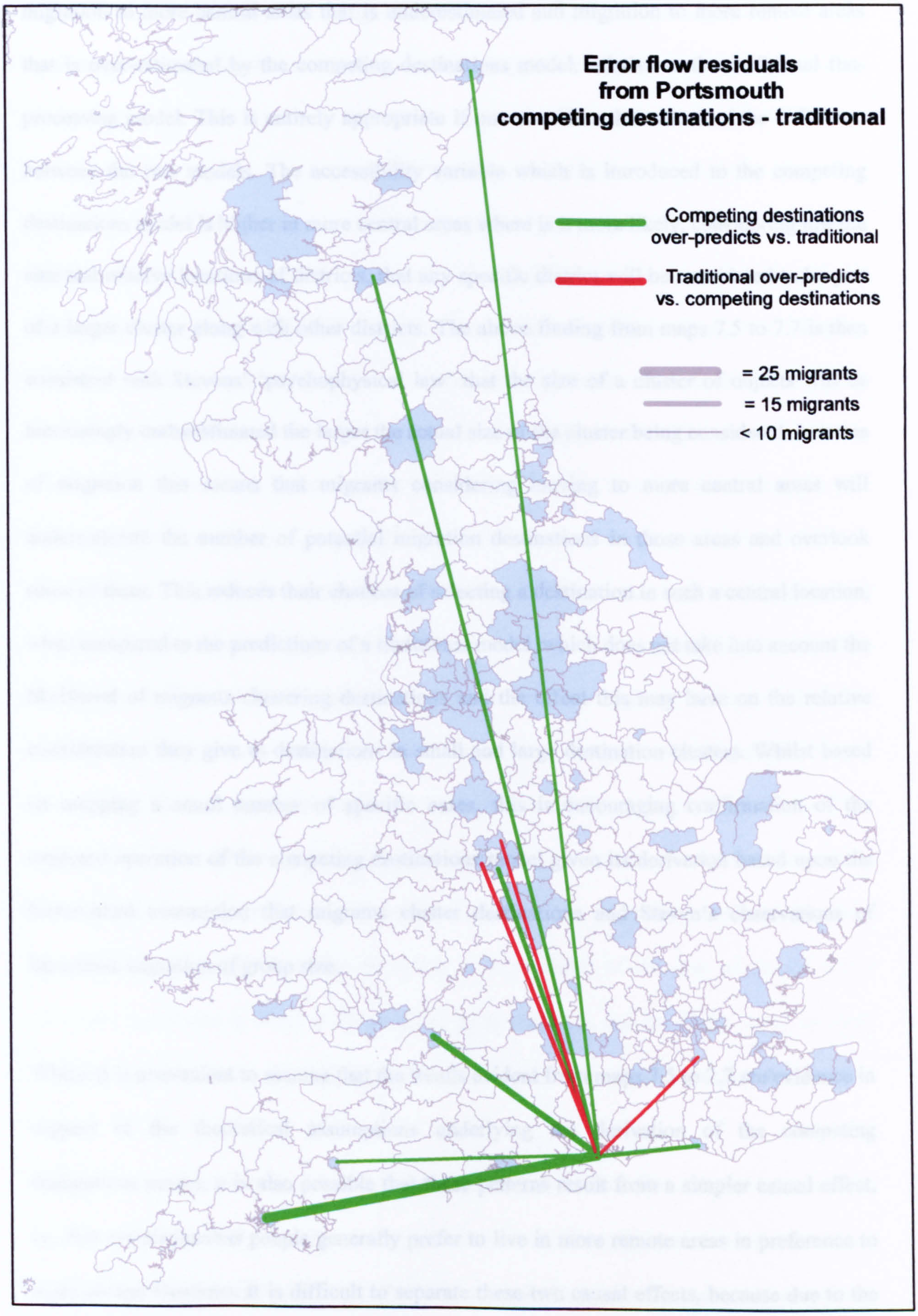
Map 7.4: Error flows residuals from Kings Lynn & W. Norfolk, competing destinations model.



Map 7.5: Residual flows from Kings Lynn and W. Norfolk, competing destinations vs. traditional.



Map 7.6: Residual flows from Derby, competing destinations vs. traditional models.



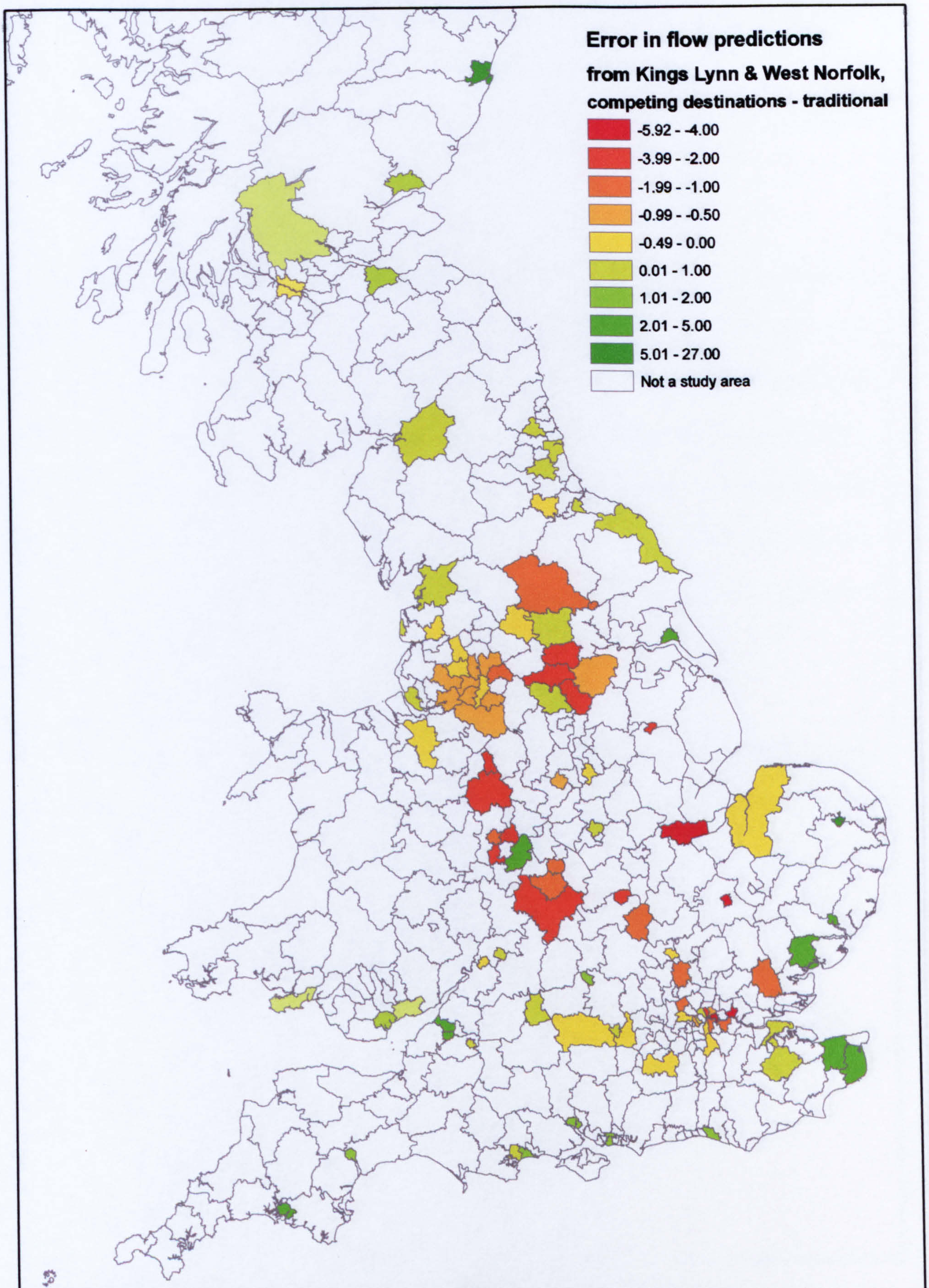
Map 7.7: Residual flows from Portsmouth, competing destinations vs. traditional models.

It is interesting to note that in all three cases presented in maps 7.5, 7.6 and 7.7, it is migration to more central areas that is underestimated and migration to more remote areas that is over-estimated by the competing destinations model, relative to the traditional flat-processing model. This is entirely appropriate if one considers that nature of the difference between the two models. The accessibility variable which is introduced to the competing destinations model is higher in more central areas where is it more likely, considering just the size and relative locations of districts, that any specific district will be considered to be part of a larger cluster along with other districts. The above finding from maps 7.5 to 7.7 is then consistent with Stevens' 'psychophysical law' that the size of a cluster of objects will be increasingly underestimated the larger the actual size of the cluster being considered. In terms of migration this means that migrants considering moving to more central areas will underestimate the number of potential migration destinations in those areas and overlook some of them. This reduces their chances of selecting a destination in such a central location, when compared to the predictions of a traditional model, which does not take into account the likelihood of migrants clustering destinations and the effect this may have on the relative consideration they give to destinations in small and large destination clusters. Whilst based on mapping a small number of specific cases, this is encouraging confirmation of the expected operation of the competing destinations model given its derivation based upon the hierarchical assumption that migrants cluster destinations and Steven's observations of inaccurate cognition of group size.

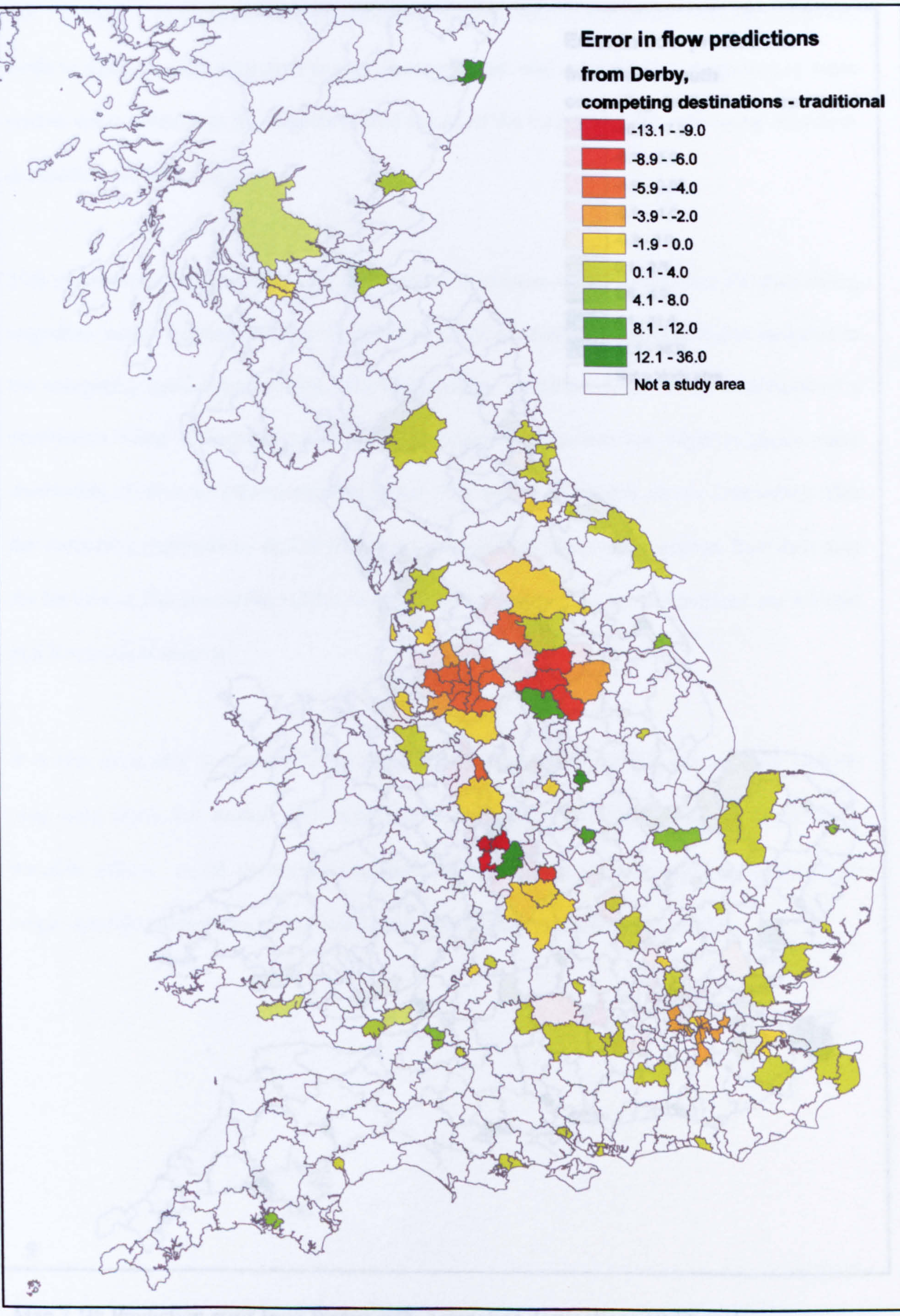
Whilst it is convenient to assume that the trends evident from maps 7.5 to 7.7 are evidence in support of the theoretical assumptions underlying the derivation of the competing destinations model, it is also possible that these patterns result from a simpler causal effect, i.e. that *ceteris paribus* people generally prefer to live in more remote areas in preference to more central locations. It is difficult to separate these two causal effects, because due to the nature of its definition, the accessibility variable, which is the only essential difference between the traditional and competing destinations models, is equally effective as a measure of the likelihood of a destination being cognized in a larger group of destinations, as it is as a measure of the centrality of a destination. Intuition suggests that both causal relationships are likely at work.

One way to determine the extent to which migrants are grouping destinations, and hence the extent to which the accessibility statistic is representing the likelihood of a particular destination being overlooked if it lies in a larger cluster of destinations, is to examine whether cluster- or regional-level characteristics are significant determinants of migration destination choice. The various nested logit models of migration destination choice applied in this research all include a regional utility variable, as they are based on the assumption that the destination choice process is simplified by the selection of a region prior to the selection of a specific destination. The results from calibrations of these nested logit models, presented in chapter 8, demonstrate that given sufficient sample size (i.e. when modelling the migration of all migrants aged 16 and over), parameter estimates for regional utility variables are statistically different from zero (at 95% confidence level) for 60% of origin-specific model calibrations. This suggests that regions, which are essentially clusters of destinations, are playing a significant role in many migrants' destination choice processes. That very likely means that at least a significant portion of the effects of the competing destination model's accessibility variable results from its action as a predictor of the size of cluster that a particular destination is likely to be cognized in by migrants, and the likely under- or over-consideration that cluster size is likely to have on overall migration levels to that destination.

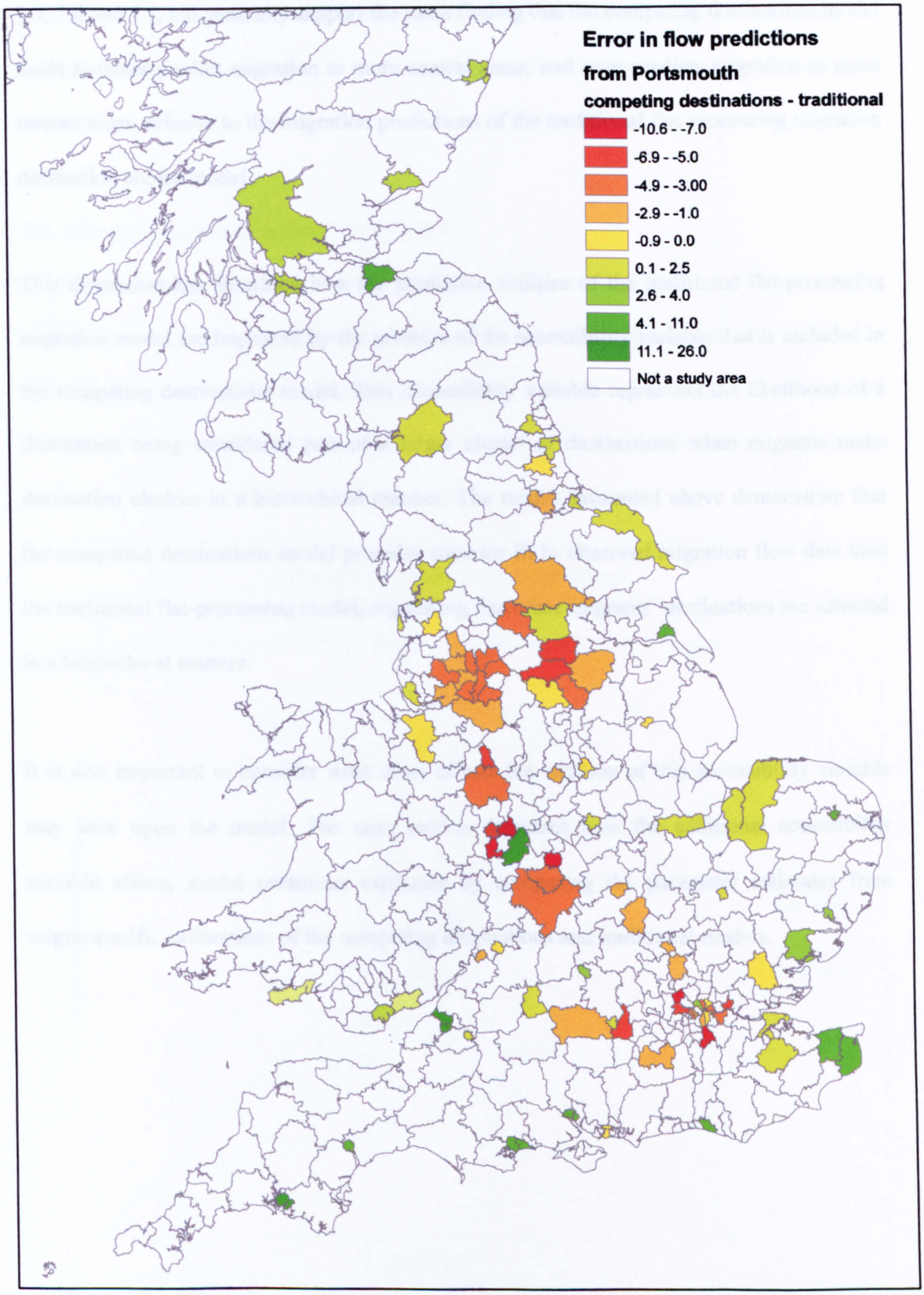
The patterns shown on maps 7.3 to 7.7 above only consider error flows that are greater than one standard deviation from the mean prediction error. This approach is useful to reduce visual congestion on the maps themselves and to make the general trends more prominent. However, it can also be useful to view the spatial patterns in the prediction errors in the flows to all destinations, in order to confirm the trends evident in from the flow maps of outlier error residuals. In order to effectively visualize all error residuals simultaneously a colour-coded map is used in preference to a flow map, with each destination district being colour-coded to present the competing destination model's over- or under-prediction of migration from a specific origin to each of 99 destinations, compared to the traditional model. Migration prediction error is used in preference to percentage prediction error in order to avoid relatively small errors in small observed flows being over-represented. Maps 7.8, 7.9 and 7.10 present the prediction error of the competing destination model compared to the traditional model for migration from the origins: Kings Lynn and West Norfolk, Derby and Portsmouth, respectively.



Map 7.8: Prediction error from Kings Lynn & W.Norfolk, competing destinations-traditional.



Map 7.9: Prediction error from Derby, competing destinations-traditional.



Map 7.10: Prediction error from Portsmouth, competing destinations-traditional.

Maps 7.8, 7.9 and 7.10 show a weaker trend than the outlier flow residuals presented in maps 7.5, 7.6 and 7.7, but generally support the same finding that the competing destinations model tends to under-predict migration to more central areas, and over-predicts migration to more remote areas, relative to the migration predictions of the traditional flat-processing migration destination choice model.

This discussion has described how the predictive abilities of the traditional flat-processing migration model are improved by the addition of the accessibility variable that is included in the competing destinations model. This accessibility variable represents the likelihood of a destination being considered part of a larger cluster of destinations when migrants make destination choices in a hierarchical manner. The results presented above demonstrate that the competing destinations model provides superior fit to observed migration flow data than the traditional flat-processing model, suggesting that some migrants' destinations are selected in a hierarchical manner.

It is also important to consider what other effects the addition of this accessibility variable may have upon the model. The next section discusses how the additional accessibility variable affects model parameter estimates by comparing the parameter estimates from origin-specific calibrations of the competing destinations and traditional models.

Differences in Parameter Estimates

This section examines how the addition of the accessibility variable to the traditional migration model to produce the competing destinations model affects the parameter estimates of the other explanatory variables.

The effects of adding the accessibility variable

The parameter estimates for each explanatory variable generated from the calibration of the traditional and competing destinations models for the 100 selected migration origins are plotted against each other in figures 7.10 to 7.15 below. The parameter estimates (and goodness-of-fit statistics) from calibrations of the traditional and competing destinations models are tabulated in full in appendices D and E, respectively.

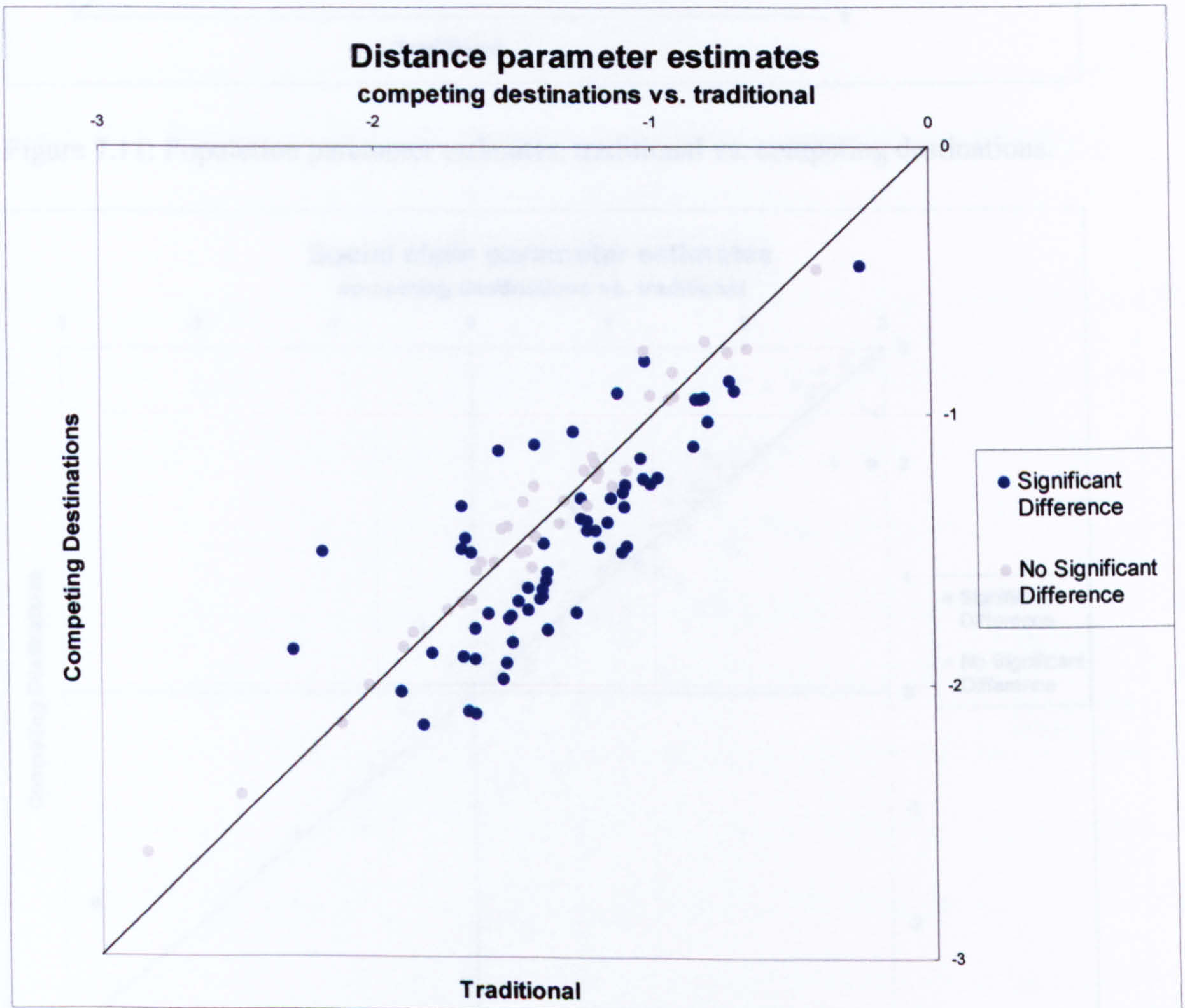


Figure 7.10: Distance parameter estimates, traditional vs. competing destinations.

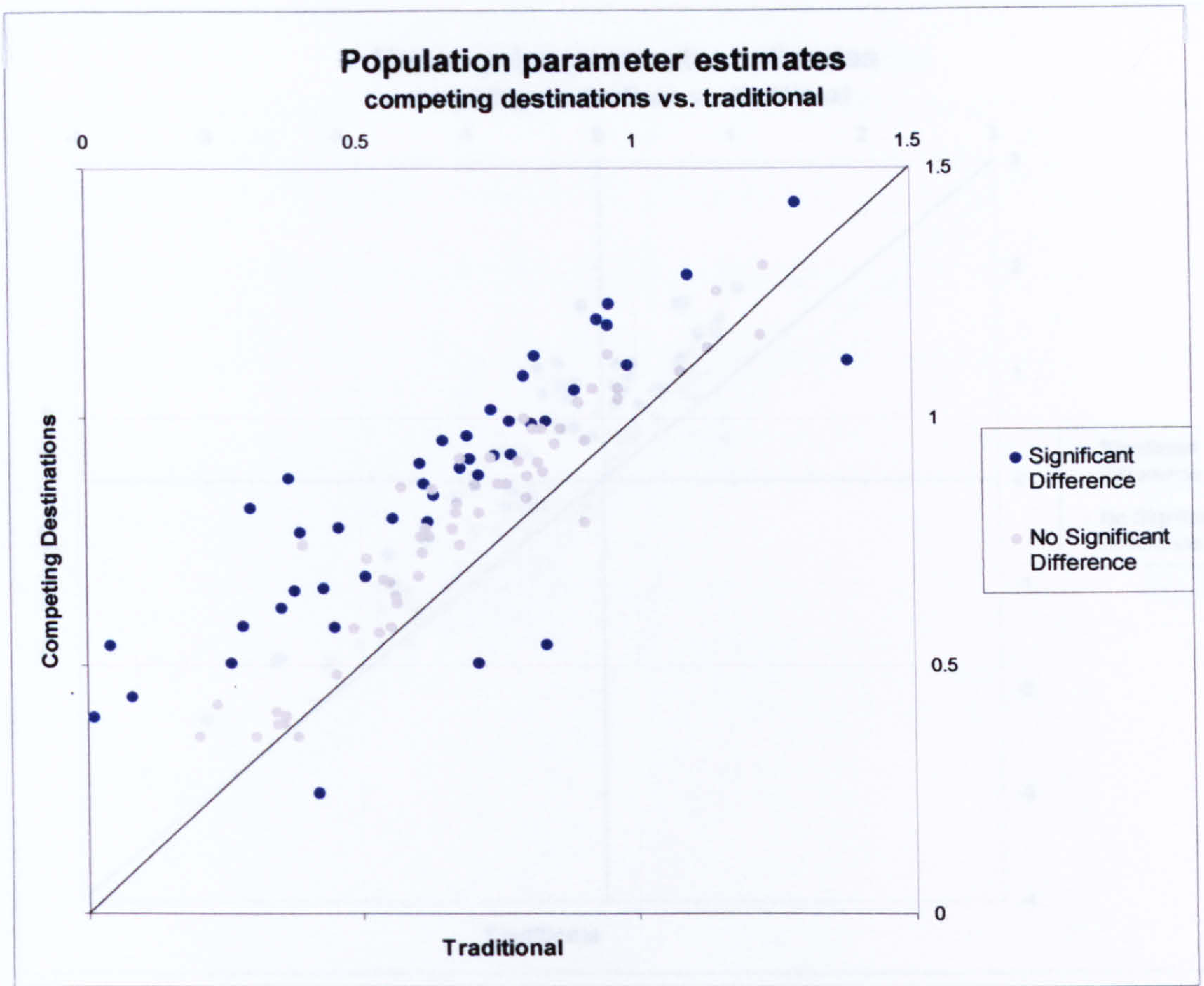


Figure 7.11: Population parameter estimates, traditional vs. competing destinations.

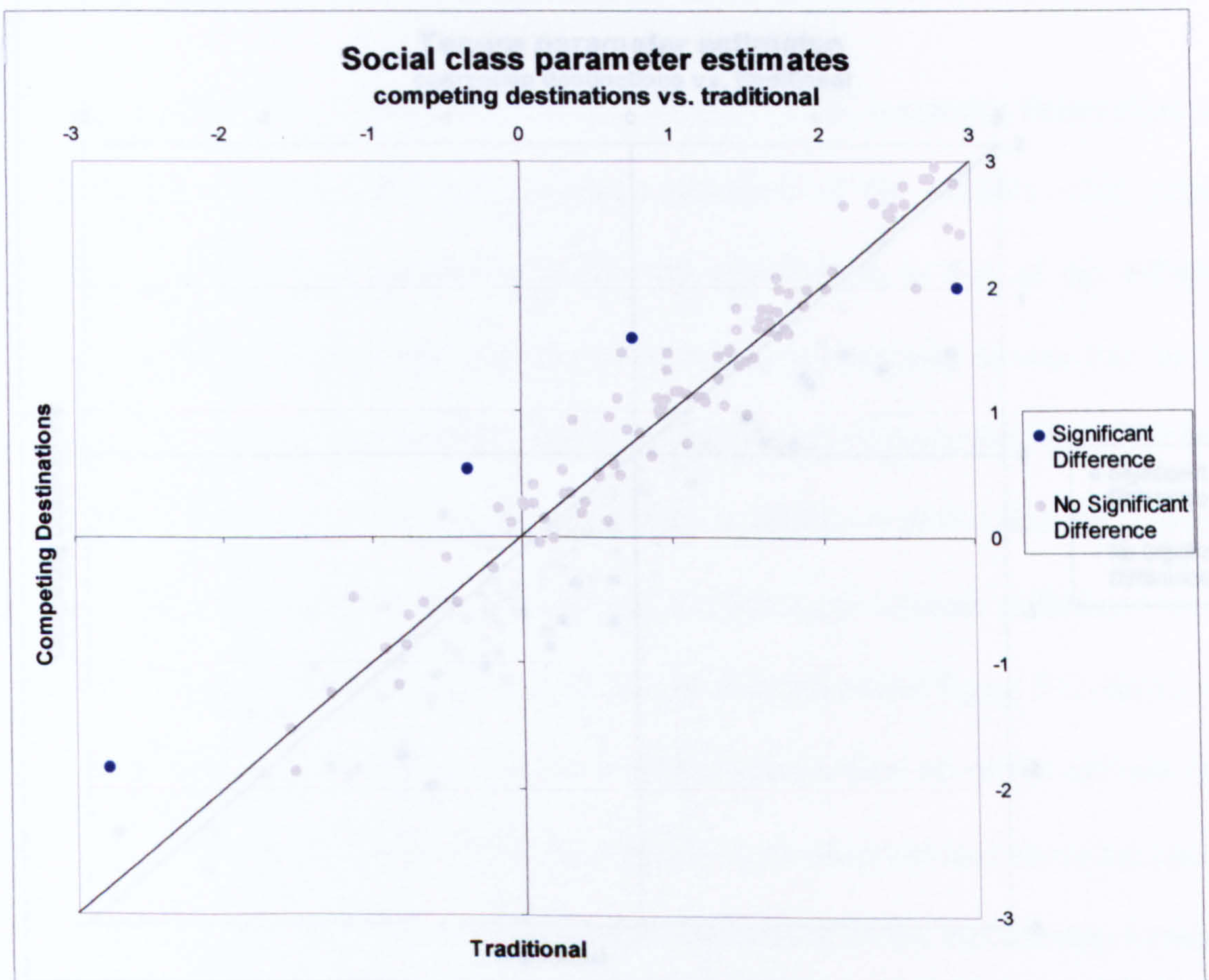


Figure 7.12: Social class parameter estimates, traditional vs. competing destinations.

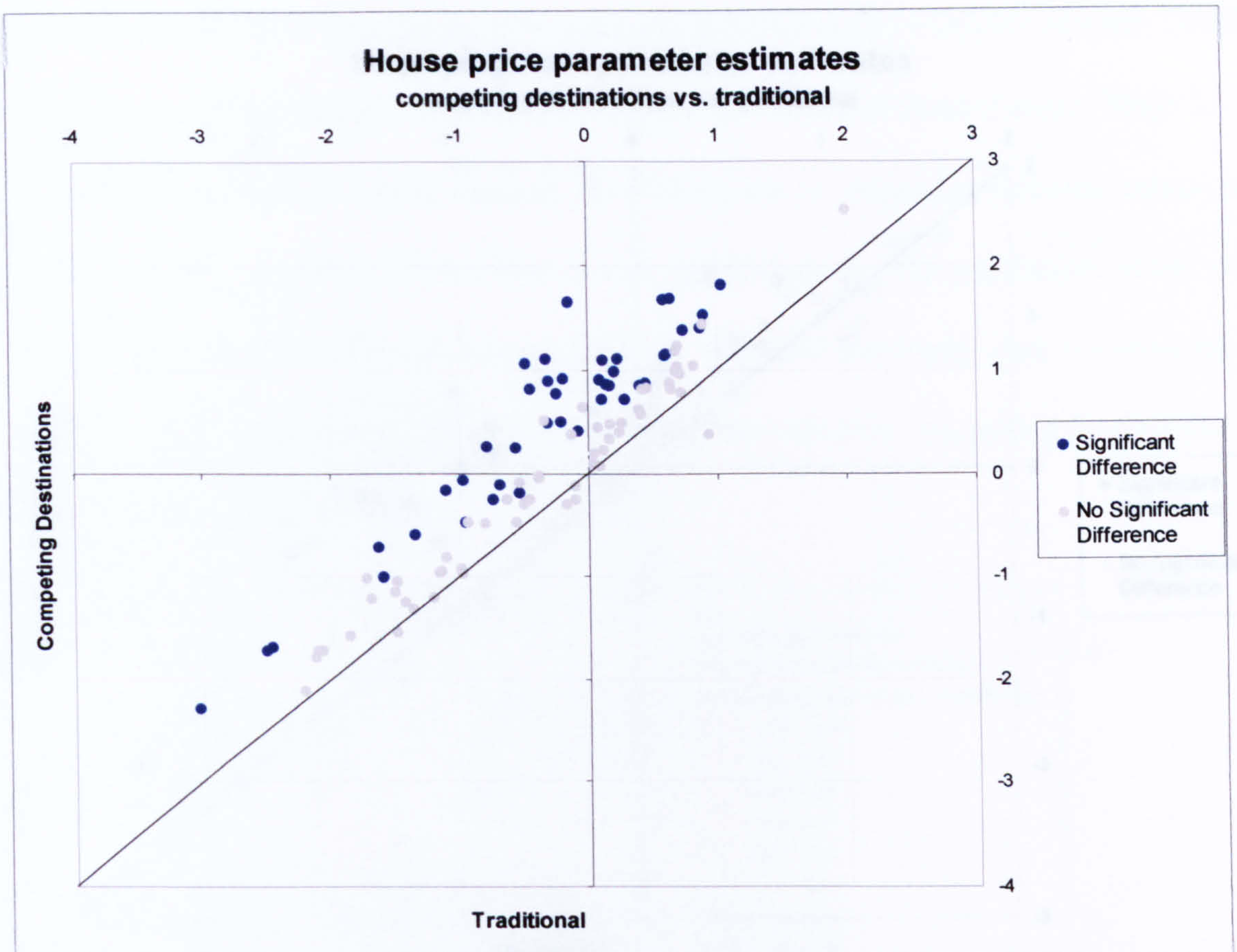


Figure 7.13: House price parameter estimates, traditional vs. competing destinations.

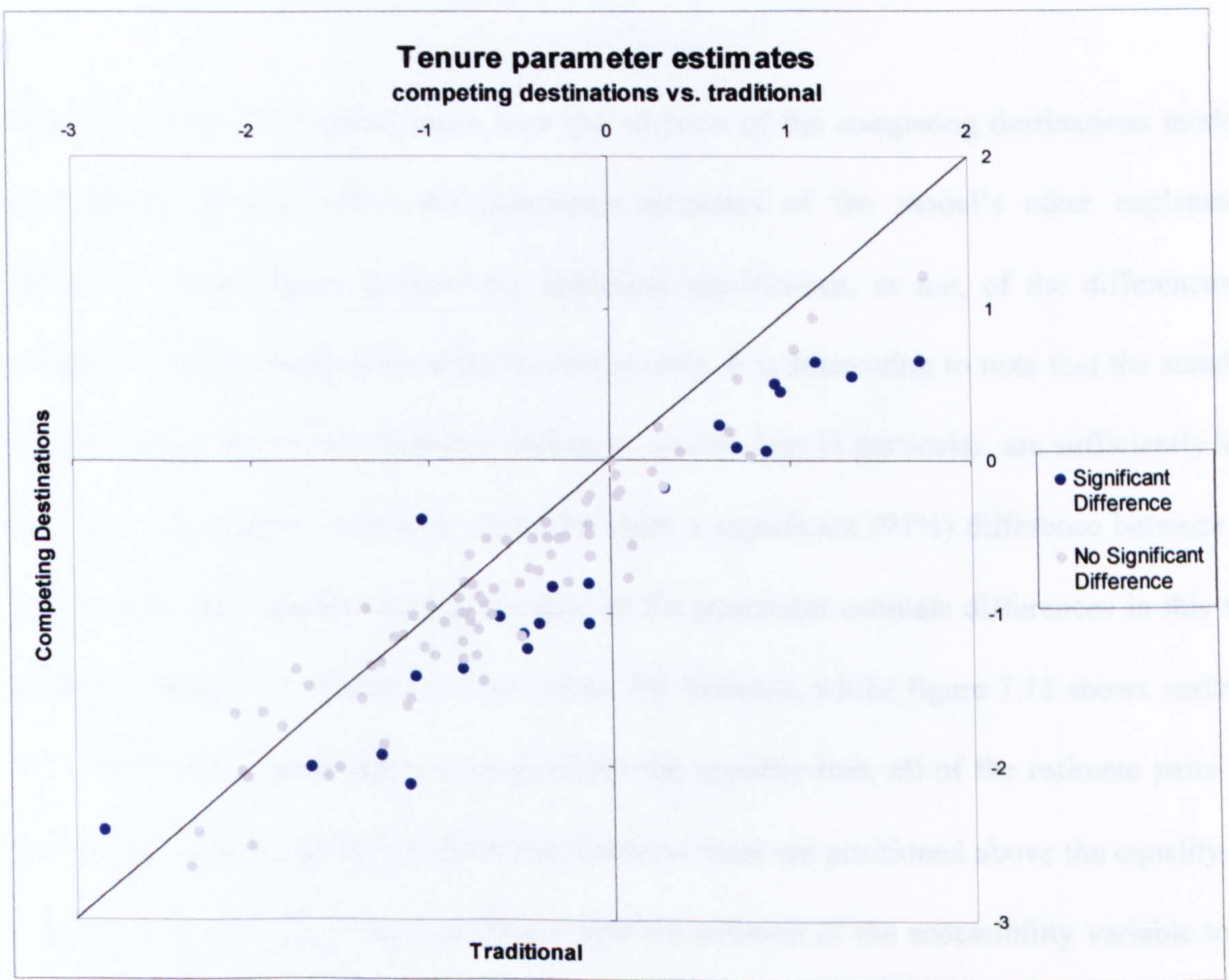


Figure 7.14: Tenure parameter estimates, traditional vs. competing destinations.

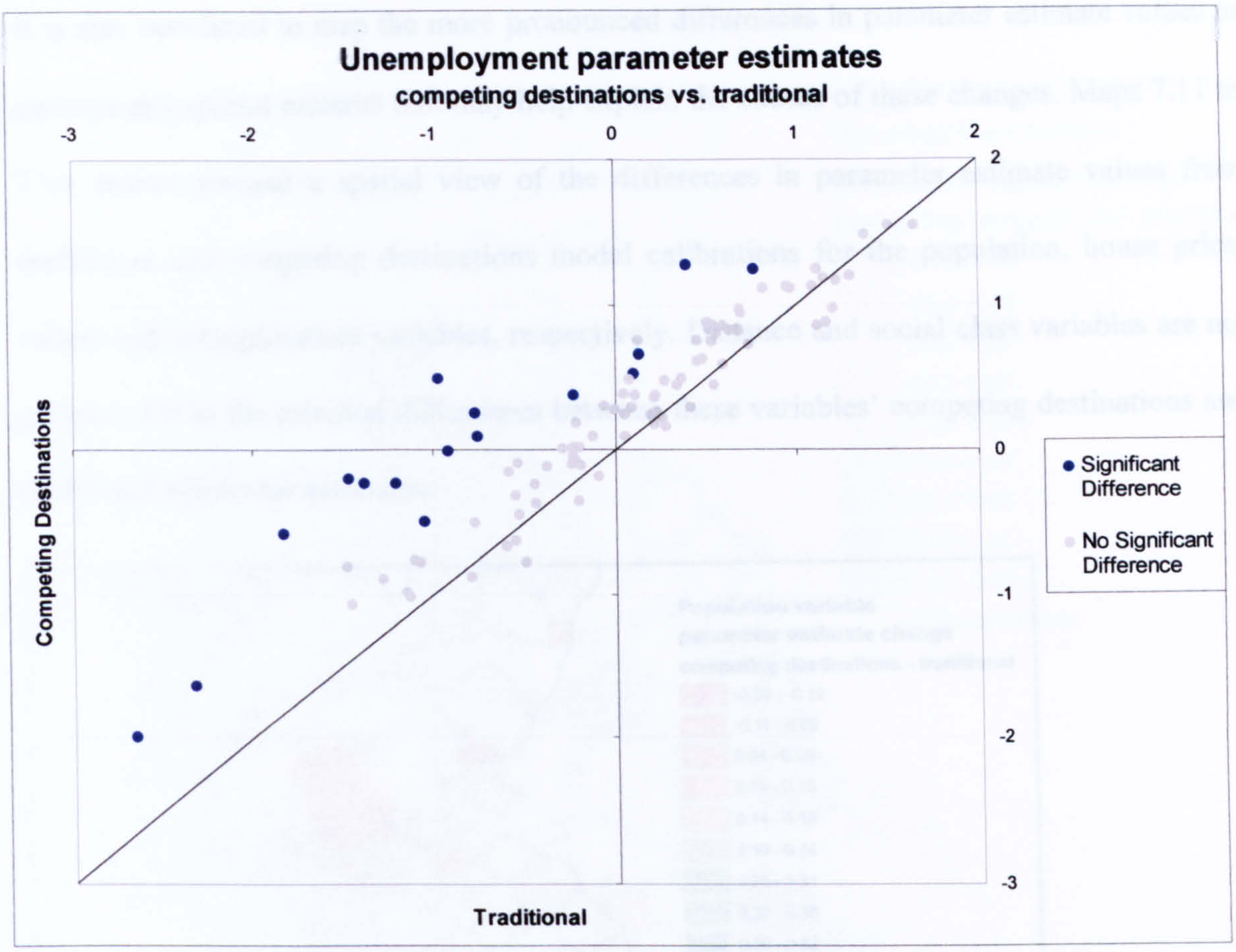
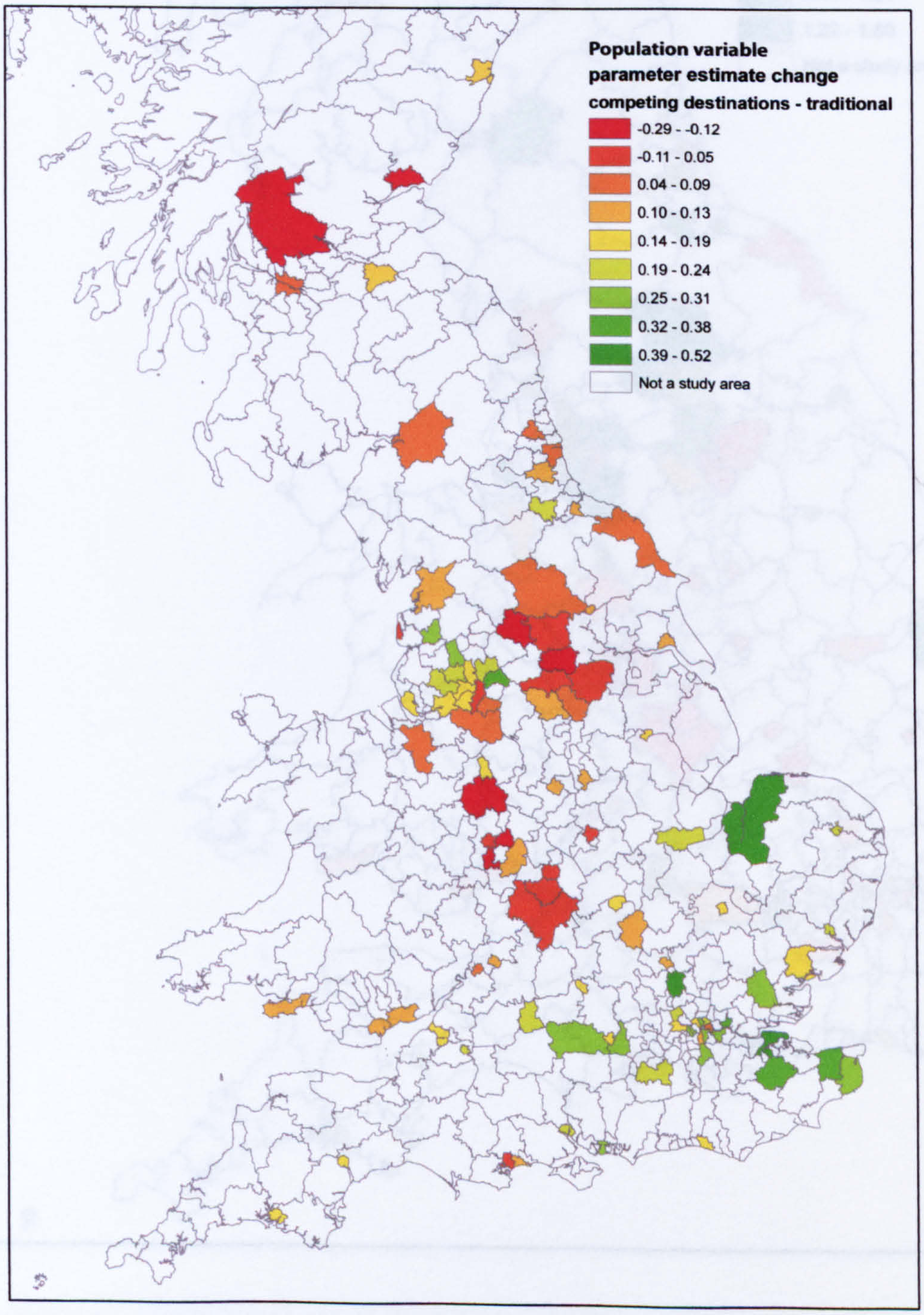


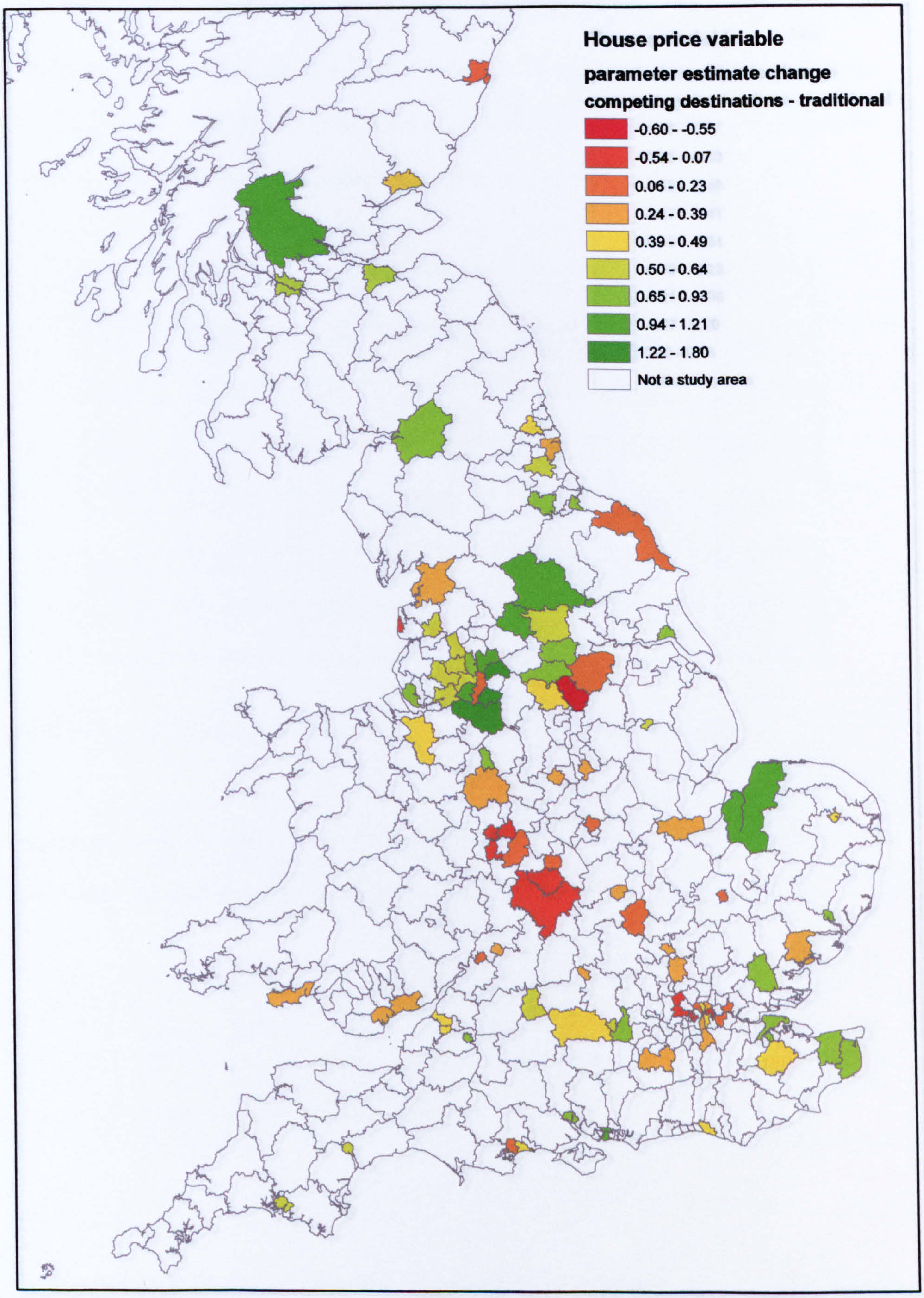
Figure 7.15: Traditional vs. competing destinations unemployment parameter estimates.

Figures 7.10 to 7.15, above, show how the addition of the competing destinations model's accessibility variable alters the parameter estimates of the model's other explanatory variables. These figures indicate the statistical significance, or not, of the differences in parameter estimate values between the two models. It is interesting to note that the standard errors of some variables' parameter estimates, social class in particular, are sufficiently high that very few origins' parameter estimates show a significant (95%) difference between the two models. Highlighting the significance of the parameter estimate differences in this way clarifies the patterns inherent in these plots. For instance, whilst figure 7.15 shows variation of unemployment parameter estimates about the equality line, all of the estimate pairs that have a statistically significant difference between them are positioned above the equality line – which adds strength to the conclusion that the addition of the accessibility variable to the model causes an increase in the values of unemployment parameter estimates.

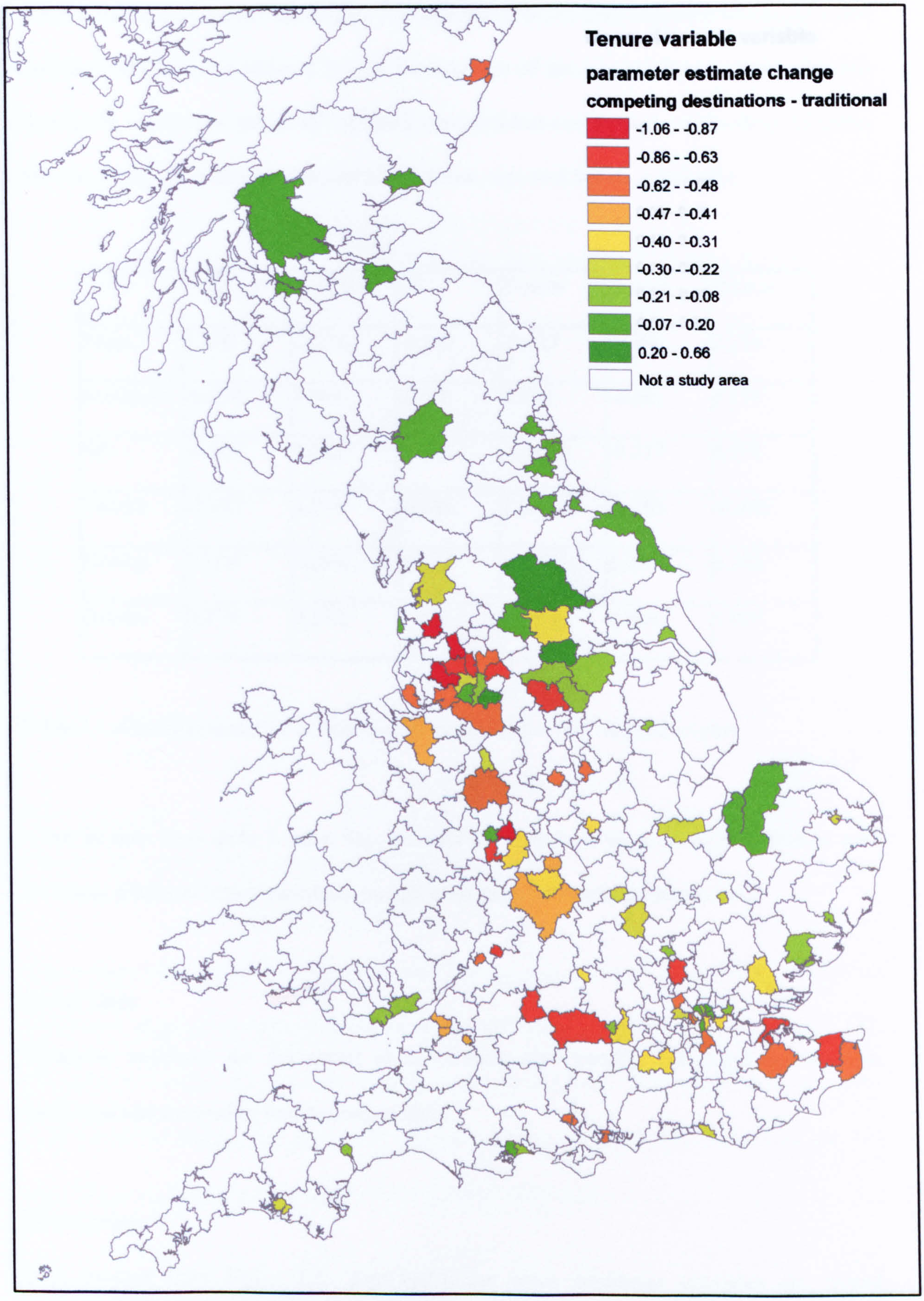
It is also beneficial to map the more pronounced differences in parameter estimate values to uncover any spatial patterns that may help explain the causes of these changes. Maps 7.11 to 7.14, below, present a spatial view of the differences in parameter estimate values from traditional and competing destinations model calibrations for the population, house price, tenure and unemployment variables, respectively. Distance and social class variables are not mapped due to the minimal differences between these variables' competing destinations and traditional parameter estimates.



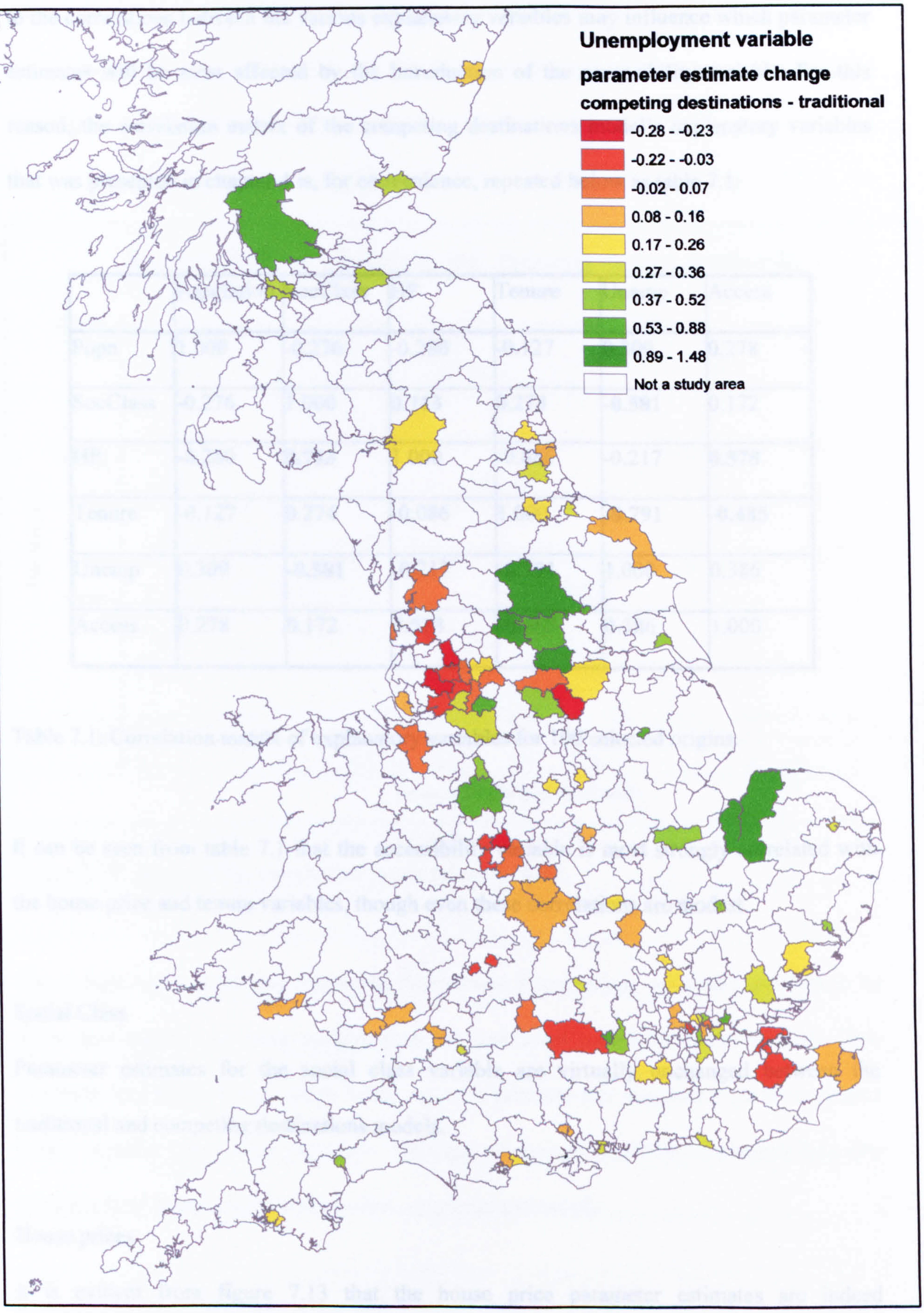
Map 7.11: Difference in population para. estimates, competing destinations – traditional.



Map 7.12: Difference in house price para. estimates, competing destinations – traditional.



Map 7.13: Difference in tenure parameter estimates, competing destinations – traditional.



Map 7.14: Difference in unemployment para. estimates, competing destinations – traditional.

It is reasonable to expect that the multicollinearity inherent in the models' specifications due to the correlations between the various explanatory variables may influence which parameter estimates will be most affected by the introduction of the accessibility variable. For this reason, the correlation matrix of the competing destinations model's explanatory variables that was presented in chapter 4 is, for convenience, repeated below as table 7.1:

	Population	SocClass	HP	Tenure	Unemp	Access
Popn	1.000	-0.276	-0.280	-0.127	0.309	0.278
SocClass	-0.276	1.000	**0.713**	0.274	**-0.581**	0.172
HP	-0.280	**0.713**	1.000	-0.086	-0.217	**0.578**
Tenure	-0.127	0.274	-0.086	1.000	**-0.791**	**-0.485**
Unemp	0.309	**-0.581**	-0.217	**-0.791**	1.000	0.386
Access	0.278	0.172	**0.578**	**-0.485**	0.386	1.000

Table 7.1: Correlation matrix of explanatory variables for 100 selected origins.

It can be seen from table 7.1 that the accessibility variable is most strongly correlated with the house price and tenure variables, though even these correlations are modest.

Social Class

Parameter estimates for the social class variable are virtually unchanged between the traditional and competing destinations models.

House prices

It is evident from figure 7.13 that the house price parameter estimates are indeed systematically affected by the addition of the accessibility variable. Indeed, comparison with plots of the other parameter estimates shows that the house price variable's parameter estimates are the most strongly affected of all the variables' estimates. It should also be noted that house prices are the most strongly correlated variable with the accessibility variable, though their correlation coefficient of 0.578 is still not a strong relationship, as can be seen from figure 7.16, below, which plots the two variables.

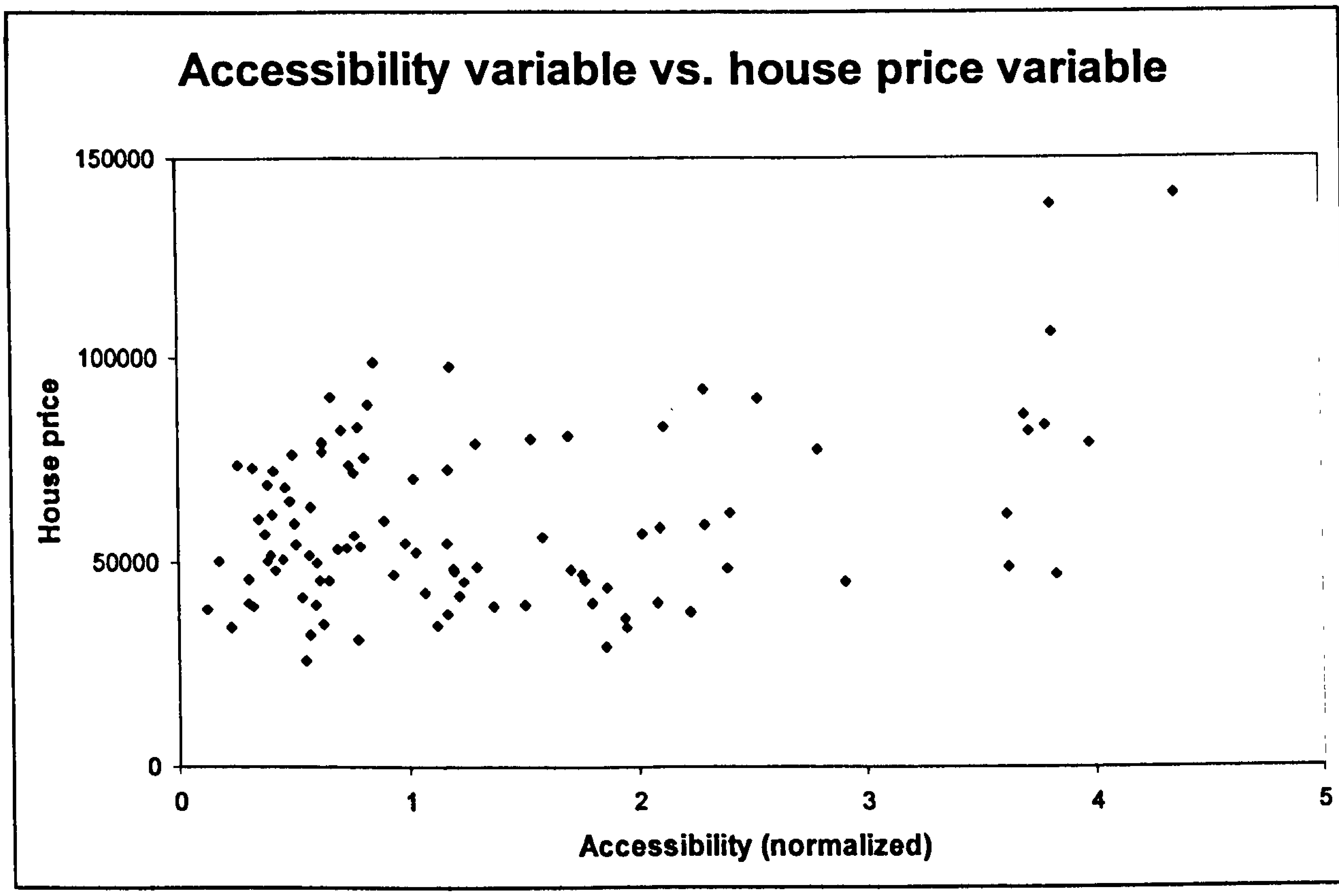


Figure 7.16: Accessibility variable vs. house prices variable.

The general trend is for the competing destinations house prices parameter estimate to be either less negative or more positive than the traditional model parameter estimate. This translates into higher house prices being less of a deterrent (for migrants from those origins with a negative house price parameter estimate), or more of an attractive characteristic (for migrants from those origins with positive parameter estimates).

To understand the observed pattern of parameter estimate changes resulting from the addition of the accessibility variable we should recall that the value of the house price variable's parameter estimate is an indication of the extent to which variation in house prices for the various possible destinations affects the level of migration into those areas *taking into account variation in all other explanatory variables*. Thus, following the addition of the accessibility variable to the traditional model, the competing destinations house price parameter estimate is describing the deterrent or attractive effect of only that portion of the variation in house prices that is not correlated with, and therefore essentially already described by the accessibility variable.

Because of the moderate positive correlation between the two variables, the negative parameter estimates for the competing destinations model's accessibility variable are essentially capturing some portion of migrants' sensitivity to higher house prices, hence the general upwards shift in the values of competing destinations model's house price parameter estimates, which suggests a reduced deterrence from areas of higher house prices.

The same argument can be phrased differently for those migration origins where traditional model house price parameter estimates are above zero. In such areas, migrants are attracted to areas with higher house prices, and the introduction of the accessibility variable with its negative parameter estimates would cause reductions in predicted migration to higher priced areas (due to the moderate correlation between accessibility and house prices) unless there were a counter-balancing increase in house price parameter estimates, to keep the models migration predictions optimized about the observed flow data.

Whilst not directly related to the differences between the tradition and competing destinations models, it is interesting to note that the house prices variable, like tenure and unemployment, appears to operate using different mechanisms for migrants from different origins. Migrants from some origins are deterred from moving to areas of higher house prices, whilst migrants from other origins appear to have a preference for such areas.

It is also interesting to note that it is not, as one might expect, migrants from areas with lower average cost of housing that are deterred most from moving to areas with higher house prices. This pattern is evident from figure 7.17 below, which plots the competing destinations house price parameter estimates for each origin-specific model calibration vs. the value of the house price variable for each origin.

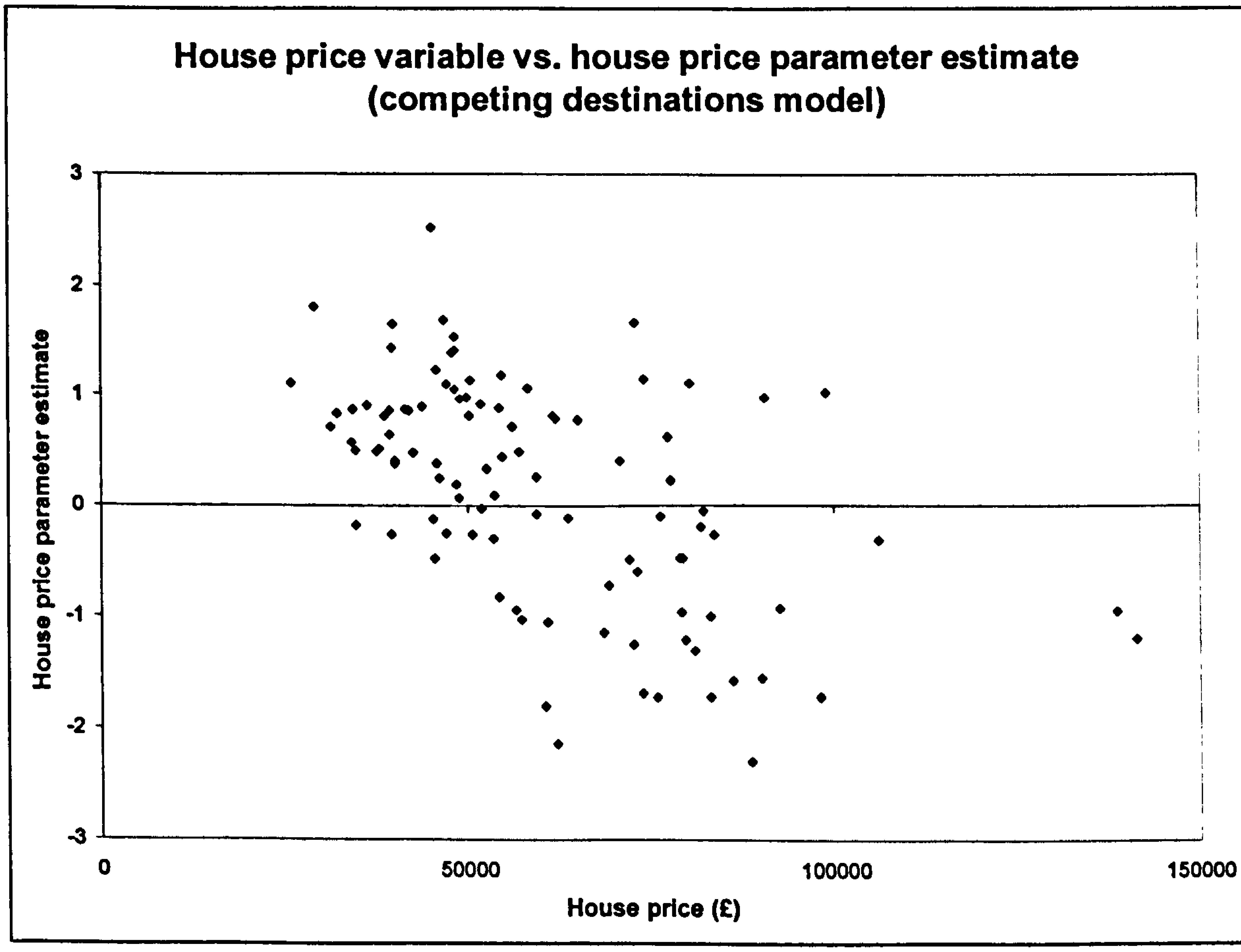


Figure 7.17: Origin house prices vs. competing destinations house price parameter estimates.

One could speculate that migrants from areas with generally cheaper housing stock, particularly younger migrants, are more likely to have to migrate to more affluent areas in order to realize ambitions of financial success, whilst there will generally be more high-paying jobs in areas of higher average house prices such that 'upward mobility' to more affluent areas is not so necessary.

Figure 7.17 also shows that migrants from areas with more expensive housing stock are generally more likely to be deterred by higher house prices. This strange finding could be caused by the effects of retirement migration which differ from most other migration in so far as many retirees will downsize from a family home to realize some capital for their retirement, or to avoid having to climb stairs, and some will also move back to their original hometown or choose to move to a traditional retirement destination such as a coastal town. It is probable that such migrants are generally less likely than most migrants to aspire to more affluent destinations.

These speculations assume differing migration behaviour by various migrant age-groups. Such variations in migration behaviour between migrant subgroups are examined in more detail in chapter 10.

Tenure and unemployment

The effects of the accessibility variable on both the tenure and unemployment parameter estimates are less obvious than the case of the house price variable, but in both cases clear trends are evident. Figure 7.14 shows that most origins' competing destinations tenure parameter estimates are lower than their corresponding traditional model parameter estimates. This pattern contrasts with the house price variable whose competing destinations parameter estimates are higher than for the traditional model. However, the fact that the correlation (albeit weak) between tenure and accessibility is a negative relationship means that it is very likely the same multicollinearity mechanism that was described above to explain changes in house prices parameter estimates, is also the cause of the generally lower tenure parameter estimates of the competing destinations model.

Unemployment parameter estimates, shown in figure 7.15, can be seen to be generally higher for the competing destinations model than for the traditional model. Again, this pattern is likely to be due to the same mechanism as described above for the house price parameter. In this case the correlation between unemployment and accessibility is positive, so the trend in the parameter estimate changes is similar to that for the house price parameter.

Social Class

The social class variable's parameter estimates conform to the intuitive reasoning that the amount of change in a variable's parameter estimates will be related to the correlation between that explanatory variable and the additional accessibility variable. The correlation between social class and accessibility is low, 0.172, and figure 7.12 shows no discernible difference between competing destinations and traditional social class parameter estimates.

Distance

The parameter estimates for the distance variables are only slightly changed by the addition of the accessibility variable. There is a minor tendency for most origins' competing destinations distance parameter estimates to be more negative than for the traditional model.

Whilst the correlation matrix presented in Table 7.1 above does not present a correlation coefficient between the distance and accessibility variables because the values of the distance variables are origin specific. However, when the distance and accessibility variables are compared on a per-origin basis, there is evidence of correlation between these variables, and the author suggests that it is the same multicollinearity problem, discussed above, that is giving rise to the slight reduction in the distance variables' parameter estimates.

Figure 7.18 and 7.19 show the relationship between the distance and accessibility variables for the origins: Aberdeen and Kensington & Chelsea.

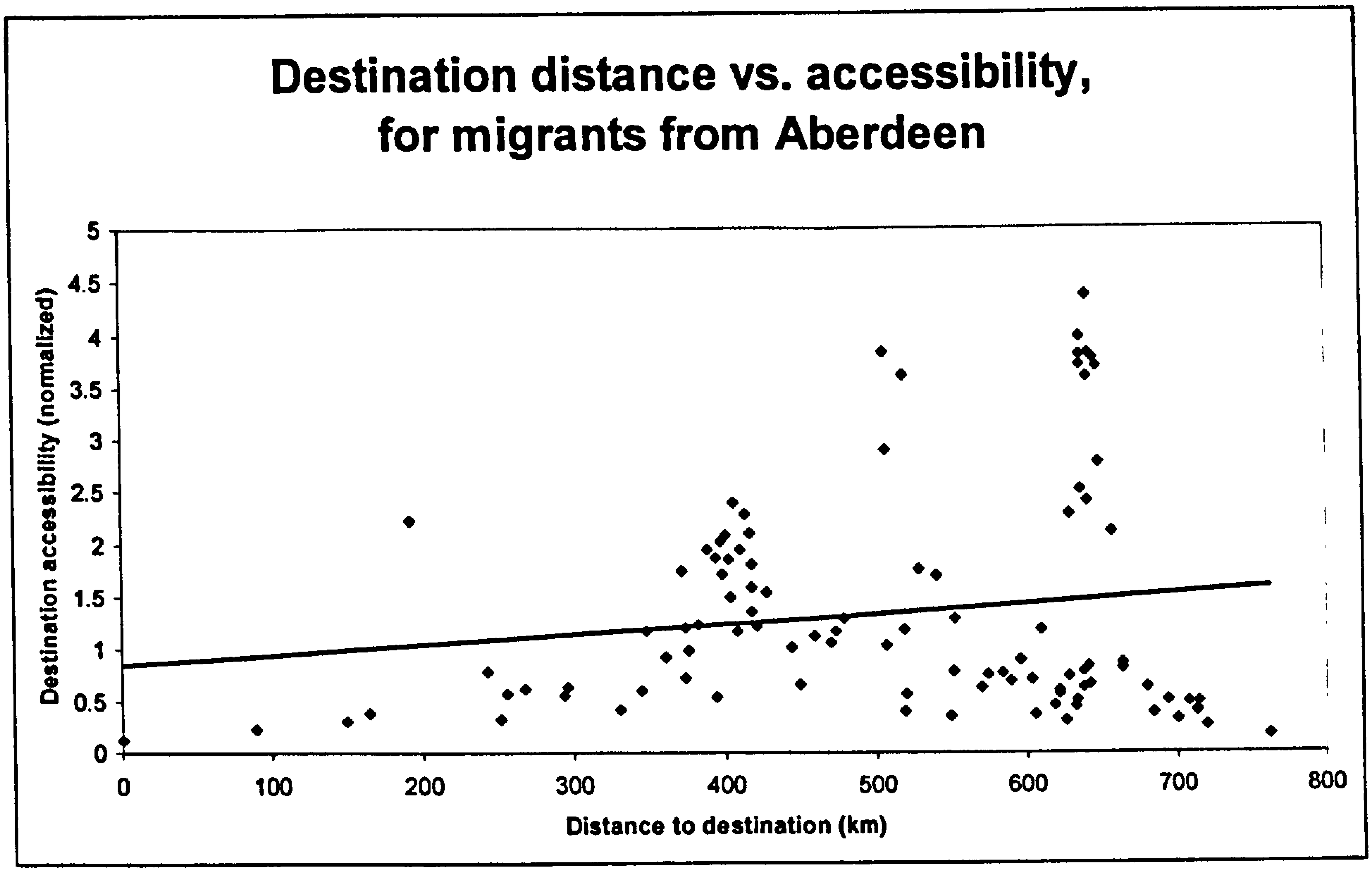


Figure 7.18: Destination distance vs. accessibility, migration from Aberdeen.

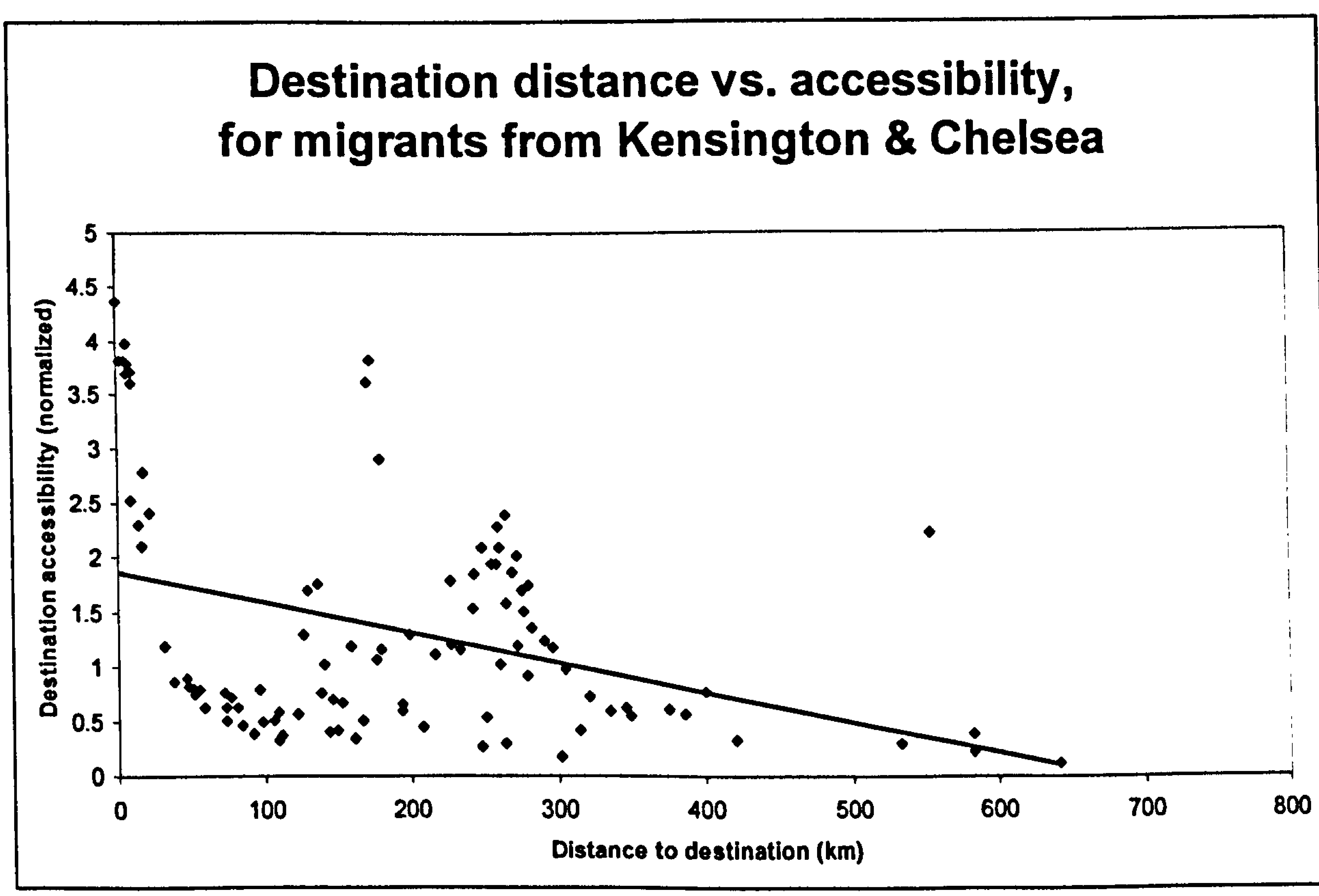


Figure 7.19: Destination distance vs. accessibility, migration from Kensington & Chelsea.

It can be seen from figures 7.18 and 7.19 that different origins experience different relationships between the accessibility and distance variables. Intuition suggests that this should be the case given that the accessibility statistic can be considered a measure of a destinations centrality as well as the likelihood of it being cognized within a larger cluster of destinations. This means that relative to a more central origin, such as Kensington and Chelsea, destinations that are more distant will generally have lower accessibility values. This matches the negative relationship between accessibility and destination distance evident in figure 7.19. Conversely, from the perspective remote origins, such as Aberdeen, the destinations with the highest accessibility will be amongst the most distant – though in this case there will also be some areas which are also 'remote' but are on the opposite end of the country, so the relationship is not so strong as for central origins. This matches the positive relationship between accessibility and destination distance evident from figure 7.18.

It is not unreasonable, given that there is some degree of correlation between the accessibility and origin-destination separation variables, to expect that the addition of the accessibility variable to the model will have some effect on the parameter estimates of the distance variable. Given the variation in the relationship between these variables for different migration origins it is not surprising to find that the accessibility variable's effect on the distance parameter estimate also varies between origins, and is also related to the origins' accessibility values. This can be seen from figure 7.20 below which plots origin accessibility against change in distance parameter estimates (competing destinations – traditional model estimate values).

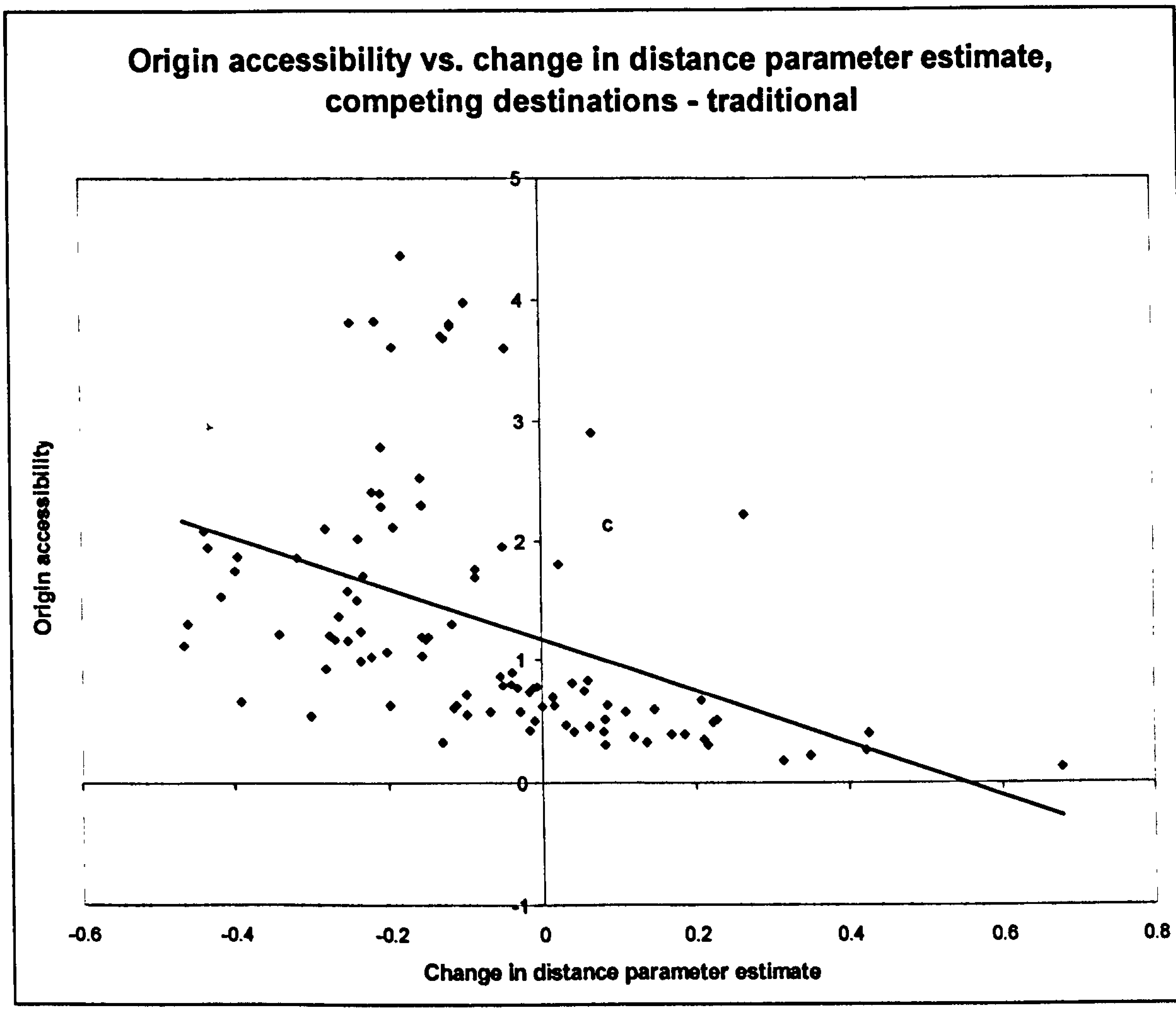


Figure 7.20: Origin accessibility vs. change (CD-trad) in distance parameter estimates

There is a very clear pattern for more central origins distance deterrence to be increased by the addition of the accessibility variable to the model (i.e. their distance parameter estimates become even more negative), whilst migrants from the most remote origins become less deterred from moving across greater distances.

This observation is consistent with the discussions of multicollinearity earlier in this chapter. Because more central origins experience a negative correlation between destination accessibility and distance, they essentially have to exhibit a stronger distance deterrence to counter the distance attractive effect of the accessibility variables - and vice versa for more remote origins.

Population

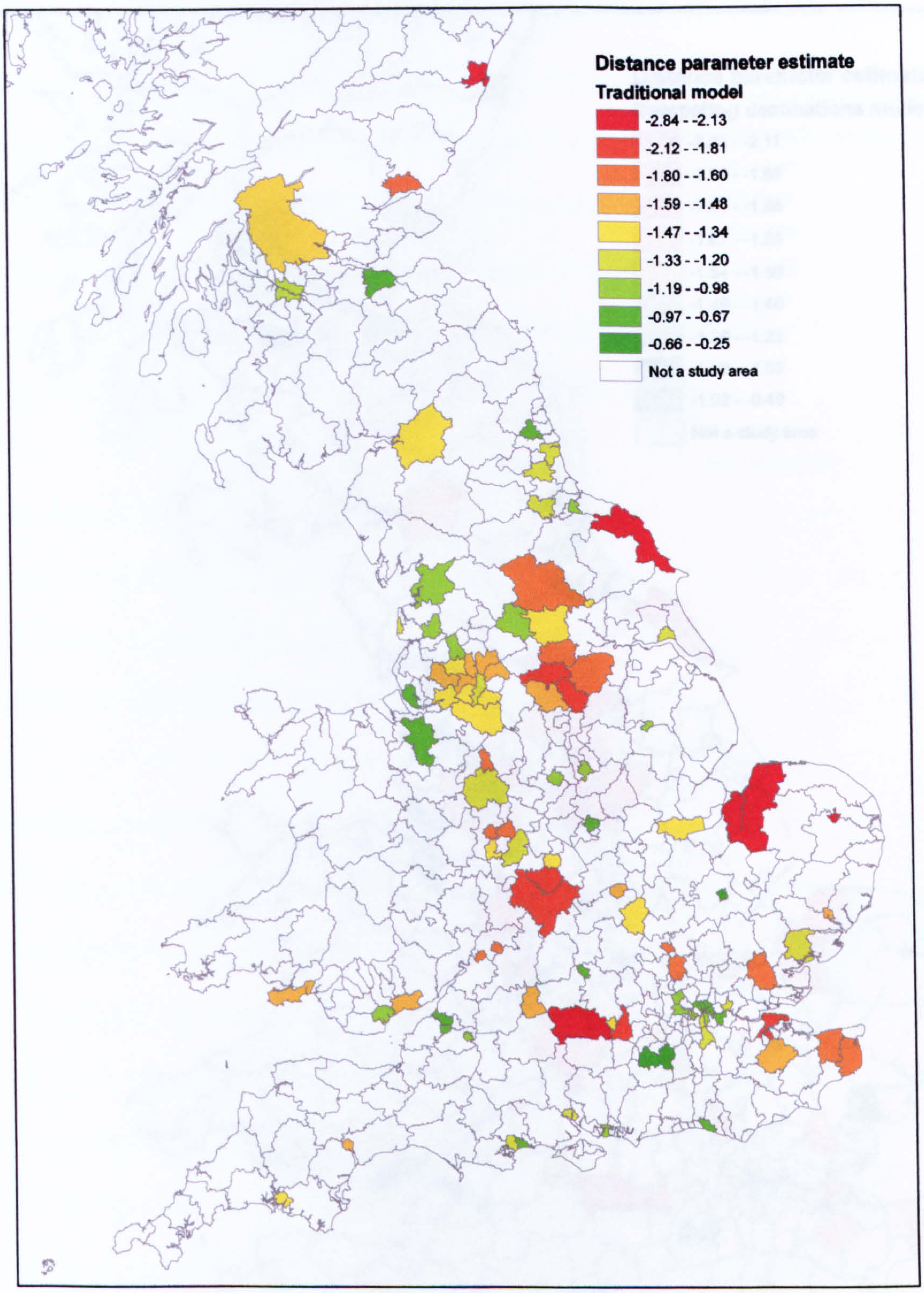
Though the correlation between the accessibility and population variables is low it is likely that the multicollinearity effect described above is also what is causing the shift in the competing destinations model's population parameter estimates. Whilst the direct correlation between the population and accessibility variables is only 0.278, comparison of maps 4.2 and 6.1 shows that they have very similar patterns of spatial variation.

In almost all calibrations of the competing destinations model the model's accessibility variable has a negative parameter estimate, meaning that the variable's net effect is to reduce predicted migration to areas with higher accessibility, (i.e. to those areas that experience more competition from neighbouring destinations due to those areas' relative positions and sizes). Because the population parameter estimate represents how variation in destination population affects migration *after controlling for variation in all other destination characteristics*, without an accessibility variable in the model the distance variable's parameter estimates are artificially lowered by the lower migration to large population areas in metro-conurbations and other high accessibility areas. Once the disinclination of migrants to move to these high accessibility areas is controlled for, by including the accessibility variable in the model, the population variable's parameter estimates will increase and will accurately reflect the attractiveness (or otherwise) of higher population as a destination characteristic.

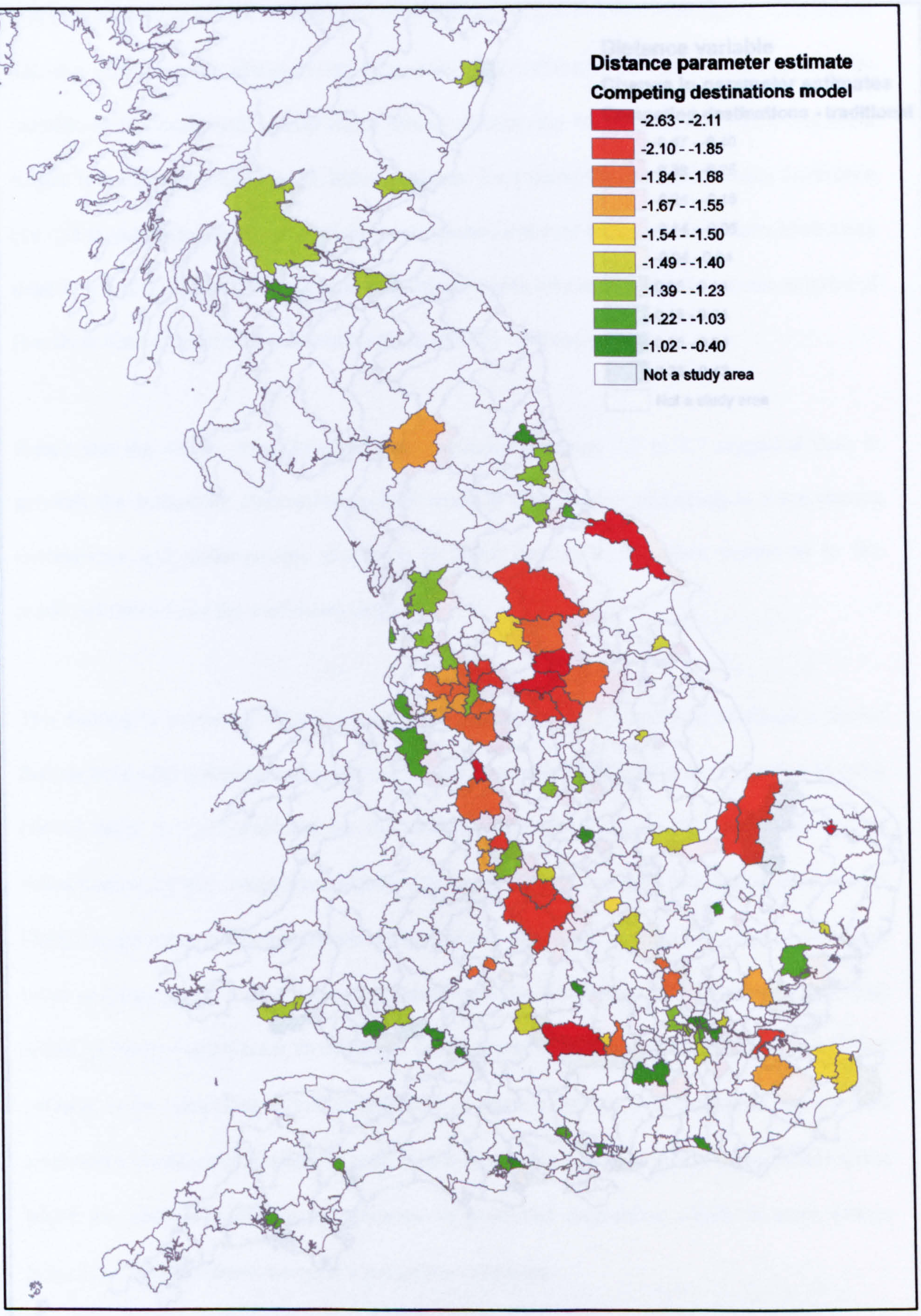
Spatial variation in parameter estimates

Mapping the parameter estimates produced by the traditional and competing destinations models is useful to illustrate the spatial variation in the influence of the explanatory variables upon migrants' destination choice behaviour. Maps 7.15 and 7.16 below show the distance variables' parameter estimates for the traditional and competing destinations models, respectively. These maps highlight the significant spatial variation in parameter estimate values.

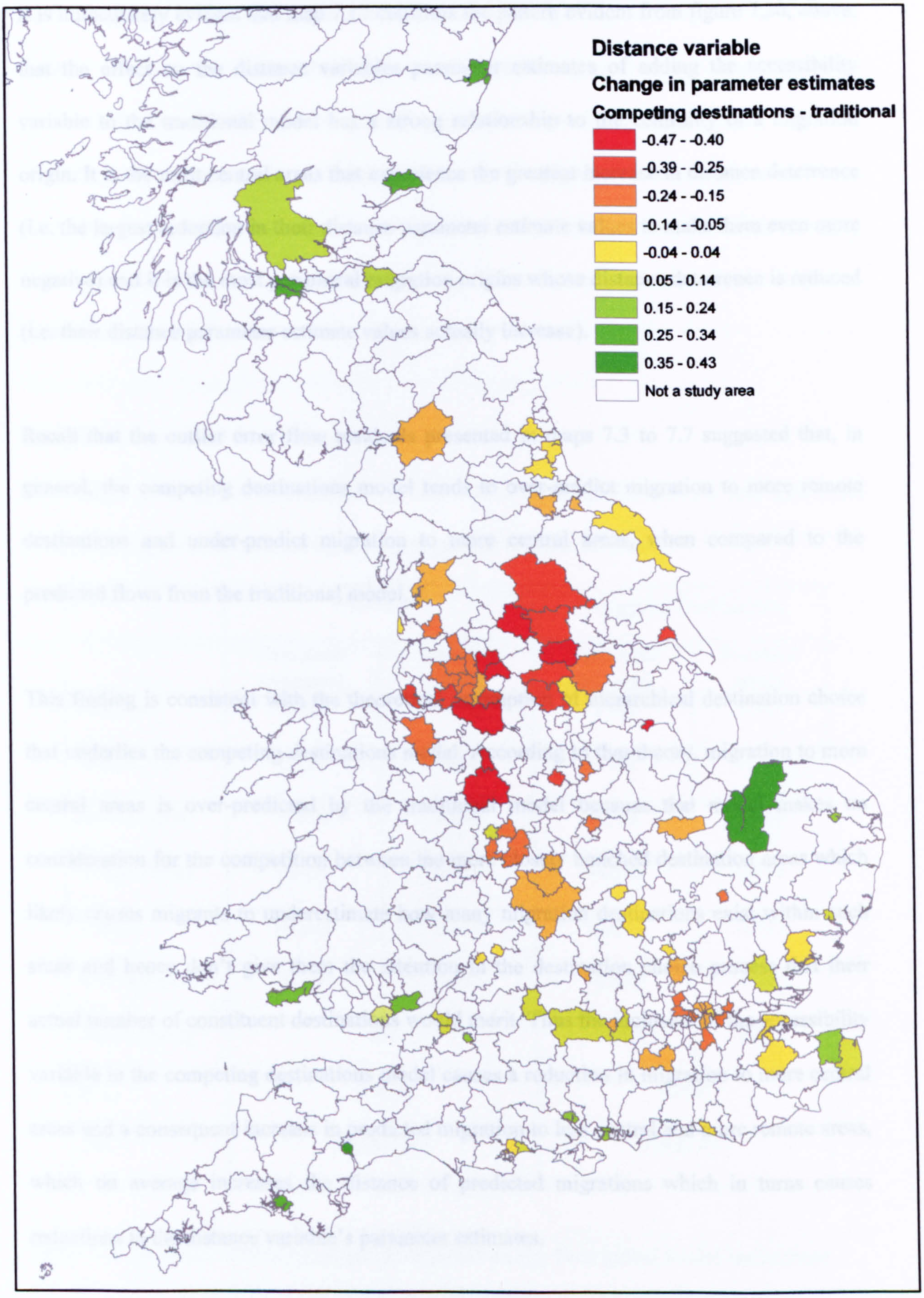
While the overall spatial patterns of the parameter estimates are similar, there are markedly more peripheral origins that exhibit lower distance deterrence (i.e. less negative distance parameter estimates) based on competing destinations parameter estimates. These patterns become clearer when the differences in the distance parameter estimates from the two models are mapped, as in map 7.17 below.



Map 7.15: Traditional model parameter estimates for the distance variable.



Map 7.16: Competing destinations model parameter estimates for the distance variable.



Map 7.17: Difference in distance parameter estimates, competing destinations vs. traditional.

It is immediately evident that map 7.17 confirms the pattern evident from figure 7.20, above, that the effect on the distance variables parameter estimates of adding the accessibility variable to the traditional model has a strong relationship to the centrality of a migration origin. It is the most central areas that experience the greatest increase in distance deterrence (i.e. the largest reduction in their distance parameter estimate values to make them even more negative) and it is the most peripheral migration origins whose distance deterrence is reduced (i.e. their distance parameter estimate values actually increase).

Recall that the outlier error flow residuals presented in maps 7.3 to 7.7 suggested that, in general, the competing destinations model tends to over-predict migration to more remote destinations and under-predict migration to more central areas, when compared to the predicted flows from the traditional model.

This finding is consistent with the theoretical assumption of hierarchical destination choice that underlies the competing destinations model. According to that theory, migration to more central areas is over-predicted by the traditional model because that model makes no consideration for the competition between the more closely bunched destination areas which likely causes migrants to underestimate how many migration destinations exist within such areas and hence don't give them the attention in the destination choice process that their actual number of constituent destinations would merit. Thus the inclusion of the accessibility variable in the competing destinations model causes a reduction in migration to more central areas and a consequent increase in predicted migration to less central and more remote areas, which on average increases the distance of predicted migrations which in turns causes reductions in the distance variable's parameter estimates.

One would expect that the finding above, that the accessibility variable increases the distance deterrence exhibited by migrants from more central origins' more than those from more remote origins, would have the effect of countering or dampening the competing destinations model's tendency to over-predict migration to more peripheral areas, especially when considering migration from more central areas. The fact that predicted flows from the competing destinations model do appear to over-predict migration to more remote locations, despite the increased distance deterrent effect discussed above, is testament to how strongly the spatial structure effects captured by the accessibility variable impact migration destination choice selection.

Global versus Local Modelling

The results from global calibrations of both the competing destinations and traditional models of migration destination choice are presented in table 7.2 below.

	Traditional		Competing Destinations	
	Parameter estimates	Standard errors	Parameter estimates	Standard errors
Distance	-1.236	0.002	-1.302	0.002
Population	0.778	0.005	0.880	0.005
Social Class	0.920	0.019	0.913	0.019
House Prices	-0.496	0.013	-0.146	0.016
Tenure	-0.271	0.016	-0.424	0.015
Unemployment	-0.267	0.015	-0.079	0.015
Accessibility	n/a		-0.408	0.011
R^2_{adj}	0.810		0.830	
AIC	92587.797		91483.427	

Table 7.2: Parameter estimates values (& standard errors) from global model calibrations.

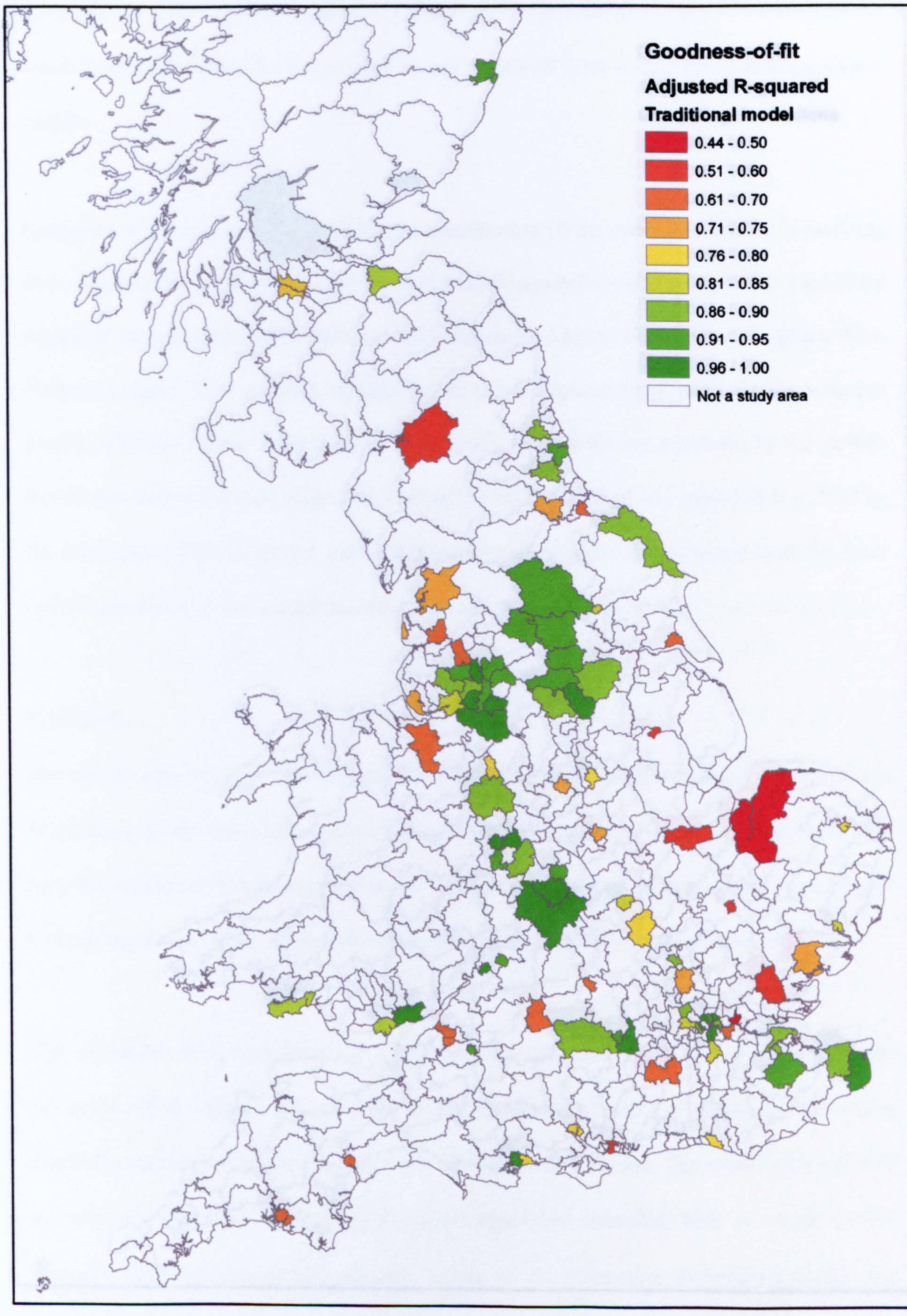
The first thing to note from these global results is that there is only a very small difference between the two models' goodness-of-fit statistics, with the competing destinations model providing a marginal 0.017 improvement in R^2_{adj} over the traditional models value of 0.810. This implies that the additional accessibility variable of the competing destinations model is having only minimal effect when the model is calibrated at a global scale. This contrasts with the origin-specific results presented above which indicate that for some origins the accessibility statistic greatly increases the model's predictive accuracy. The competing destinations models' lower AIC value confirms that the improvement in goodness-of-fit more than offsets the competing destinations model's more complex formulation (i.e. the fact that it has one more explanatory variable than the traditional model).

Of course, two models can exhibit very similar goodness-of-fit to observed migration data and yet still make very different migration predictions. The R^2_{adj} and AIC statistics don't provide any information about which specific predicted flows are closest to or furthest from observed migration flows. Thus the almost unchanged goodness-of-fit between the two global model calibrations should not be interpreted as meaning the models' predicted flows, and parameter estimates, have not changed. Indeed it is evident from table 7.2 that there have been major changes in all the parameter estimates, all of which are statistically significant changes at way above a 95% confidence level.

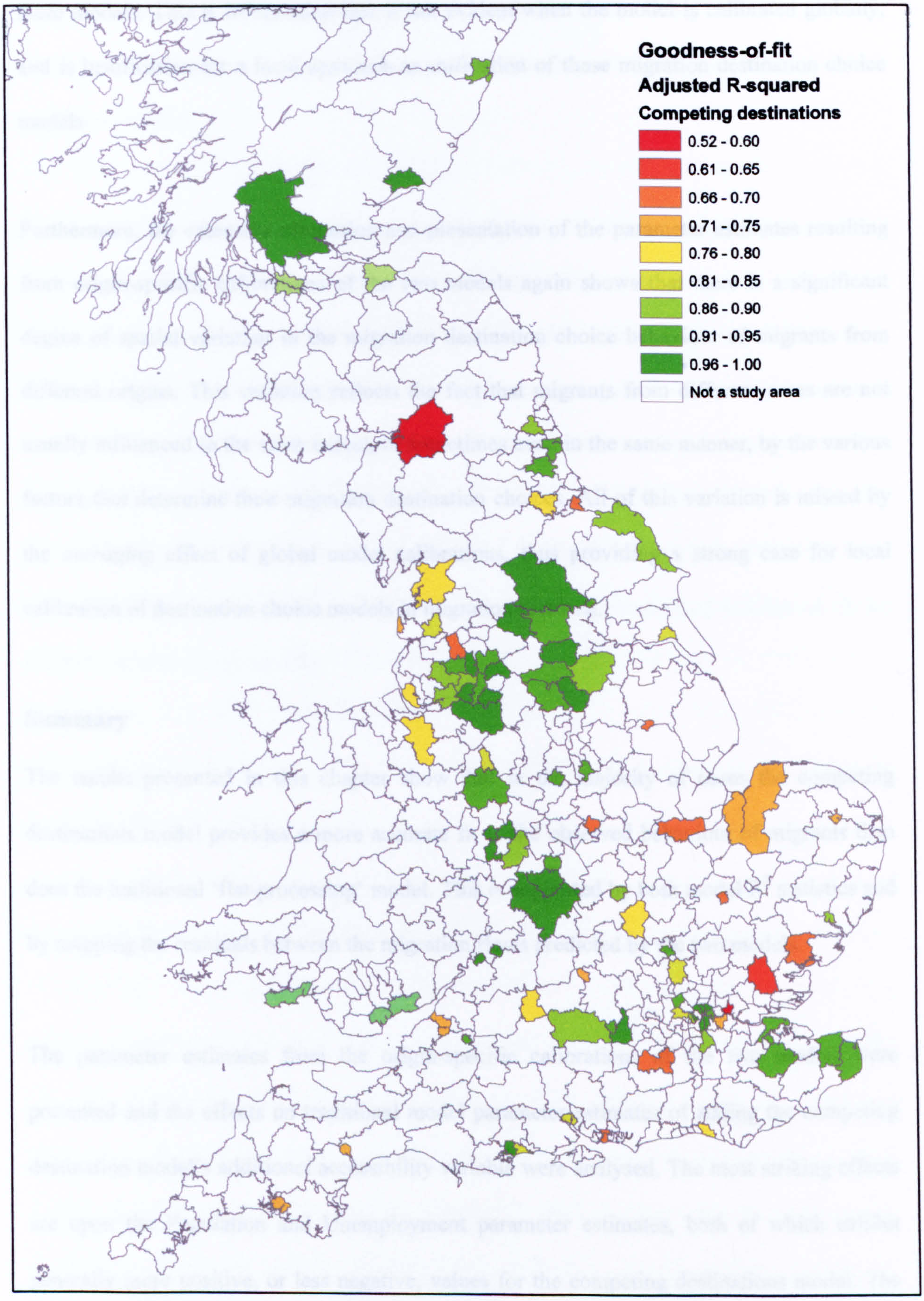
It is likely that these differences in the variables' parameter estimates, when comparing traditional and competing destinations models, are due to multicollinearity – as was discussed above in the context of origin-specific parameter estimates. Indeed, it can be seen from table 7.1 that house price, tenure and unemployment are the variables most strongly correlated with the new accessibility variable, with correlations of 0.578, -0.485 and 0.386, respectively. It is important at this point to consider that any inaccuracies arising from the multicollinearity between the accessibility and other explanatory variables should be assessed with respect to the inherent inaccuracy of the less complete model that does not contain the accessibility variable. Optimal model specification is a balance of these two factors: ensuring a sufficiently rich variety of explanatory variables, whilst maintaining acceptably low multicollinearity between these variables. Given the low number of explanatory variables in this model, and the generally improved Akaike information criterion of the more complex models, the author contends that the benefits of inclusion of the additional accessibility variable outweighs the negative effects of the multicollinearity that it introduces.

It is apparent from table 7.2 that the inclusion of the accessibility variable in the competing destinations model causes a sizeable change in the house price, tenure and unemployment parameter estimates. However, it is evident from the scatter plots of these variables' origin-specific parameter estimates (see figures 7.13, 7.14 and 7.15) that the overall patterns of change in these variables' parameter estimates are not actually as marked as the global parameter estimates suggest. For instance, the global results suggest that the effect of the house price variable is reversed from a deterrent effect to a mild attractive effect when the accessibility variable is added to the model. However, figure 7.13 shows that the competing destinations model predicts that migrants from a quarter of all origins are still deterred by higher house prices and that the traditional model predicted that migrants from almost half of the 100 origins were attracted to more expensive areas. This considerable variation is completely hidden in the results of global model calibrations.

Not only do global calibrations not permit useful comparisons of these models, they also have the disadvantage of averaging out the considerable spatial variation in the behaviour of migrants. If we look at the goodness-of-fit, as indicated by the R^2_{adj} statistic, the local calibrations show a great variation on the predictive abilities of the models for migrant leaving different origins. Maps 7.18 and 7.19, below, show the R^2_{adj} statistics from origin-specific calibrations of the traditional and competing destinations models, respectively.



Map 7.18: R^2_{adj} statistics from origin-specific calibrations of the traditional model.



Map 7.19: R^2_{adj} statistics from origin-specific competing destinations model calibrations.

These maps clearly show a significant degree of spatial variation in the goodness-of-fit of these models. This is information that is not evident when the model is calibrated globally, and is justification for a local approach to calibration of these migration destination choice models.

Furthermore, the extensive discussion and presentation of the parameter estimates resulting from origin-specific calibrations of the two models again shows that there is a significant degree of spatial variation in the migration destination choice behaviour of migrants from different origins. This variation reflects the fact that migrants from different areas are not usually influenced to the same extent, or sometimes even in the same manner, by the various factors that determine their migration destination choices. All of this variation is missed by the *averaging* effect of global model calibrations, thus providing a strong case for local calibration of destination choice models in migration research.

Summary

The results presented in this chapter show that in the majority of cases the competing destinations model provides a more accurate fit to the observed behaviour of migrants than does the traditional 'flat-processing' model. This is supported by both model R^2 statistics and by mapping the residuals between the migration flows predicted by the two models.

The parameter estimates from the origin-specific calibrations of the two models were presented and the effects on traditional model parameter estimates of adding the competing destination model's additional accessibility variable were analysed. The most striking effects are upon the Population and Unemployment parameter estimates, both of which exhibit generally more positive, or less negative, values for the competing destinations model. The parameter estimates of the Social Class variable are virtually identical for both models.

The extensive variation in both the goodness-of-fit and also the parameter estimates of the origin-specific calibrations of both models was highlighted and was presented as another example of the benefits that result from the local calibration of destination choice models in migration research.

In summary, the results from calibrating the competing destinations model presented and discussed in this chapter demonstrate significantly improved predictive ability over traditional 'flat-processing' models. This is further evidence that the hierarchical decision-making principles that underlie the derivation of the competing destinations model are a more accurate reflection of individual migrants' actual migration destination choice processes than traditional flat processing paradigms.

The next chapter discusses the results of another family of hierarchical destination choice models, the nested logit models.

Chapter Eight

Nested Logit Models

This chapter presents the results obtained from calibrating the various migration destination choice models based on nested logit discrete choice modelling. Comparisons are made between the predictive ability of these models vs. the traditional 'flat-processing' migration model. Also, the parameter estimates generated by the nested logit models and those of the traditional model are compared and contrasted.

The results from origin-specific calibrations of the discrete nested logit model are presented and discussed, and the sensitivity of the results from these calibrations to the specific discrete regionalisation against which they are calibrated is demonstrated. Reducing this sensitivity is the rationale for the development and calibration of the weighted nested logit model, a novel variation of the nested logit formulation that employs a more intuitively acceptable probabilistic regionalization when assessing the impact of regional characteristics on migration destination choice.

The results from origin-specific calibrations of the weighted nested logit model are presented and discussed with respect to the traditional model of migration destination choice. The patterns of improvement in predictive ability that this weighted nested logit model provides over the traditional model are shown to be spatially different from the patterns of improvement that the competing destinations model provides over the traditional model, (this finding is discussed in more detail in chapter 9 which compares results from various hierarchical migration models directly). This difference in the models' patterns of improvement over the traditional model suggests that the accessibility variable of the competing destinations model and the regional utility variable of the nested logit models operate in fundamentally different ways.

To explore whether this is the case a hybrid weighted nested logit model was formulated by including the accessibility variable as an additional explanatory variable alongside the regional utility variable in the weighted nested logit model. The results of origin-specific calibrations of this hybrid weighted nested logit model are also presented and discussed in this chapter.

Operationally, both the discrete and weighted nested logit models require a pre-determined regionalization as a spatial context within which they are calibrated. Because these regionalizations are generated independently for each migrant origin, the nested logit models should not really be calibrated globally, as no single regionalization is appropriate to migrants from all origins. However, in the spirit of experimentation, and in order to provide some global nested logit results (albeit somewhat suspect results) that can be compared with global results from the traditional model, global nested logit models were calibrated using discrete and probabilistic regionalizations that were generated for an origin that could be considered somewhat 'average' in its location, Derby. The results from these model calibrations are presented at the end of this chapter.

The discrete and probabilistic regionalizations that were used when producing the results presented here were generated according to the methods previously described in chapter 6. For each origin, up to 5,000,000 iterations of the regionalization algorithm were performed with the goal of producing 10,000 valid discrete regionalizations per origin. The single 'best' regionalization, defined as the one exhibiting the lowest regional information variance (see chapter 6), was then selected for use when calibrating the discrete nested logit model. A probabilistic regionalization was then constructed for each origin from the 'best' 10% of these 10,000 valid discrete regionalizations, using the method described in chapter 6.

To avoid problems arising from small sample sizes all the results presented in this chapter, as in chapter 7, relate to the migration behaviour of *all* migrants aged 16 and over. Variation in migration behaviour between age, gender and marital status migrant subgroups is examined in chapter 10.

Goodness-of-Fit

The predictive ability of the models is assessed using both the adjusted R^2_{adj} and AIC statistics introduced in the previous chapter, and also by examining error flow residuals. The R^2_{adj} statistic represents the proportion of the variance in observed migration destination choice behaviour that is explained by the regression on the explanatory variables using the generated parameter estimates. The AIC statistic is an absolute measure of the degree of variation of the predicted migration about the observed flows.

Discrete nested logit model

Comparison of the R^2_{adj} and AIC statistics for the discrete nested logit and traditional models are presented in figures 8.1 and 8.2. The tight correlation seen in figure 8.1 indicates that the discrete nested logit model provides only minor improvement in predictive ability over the traditional model for the vast majority of origins. A number of outlier origins have been identified on figures 8.1 and 8.2: Barking and Dagenham, Blackburn, Derby, Kings Lynn & West Norfolk, Plymouth, St. Albans and Wigan - these origins are labelled with their initials.

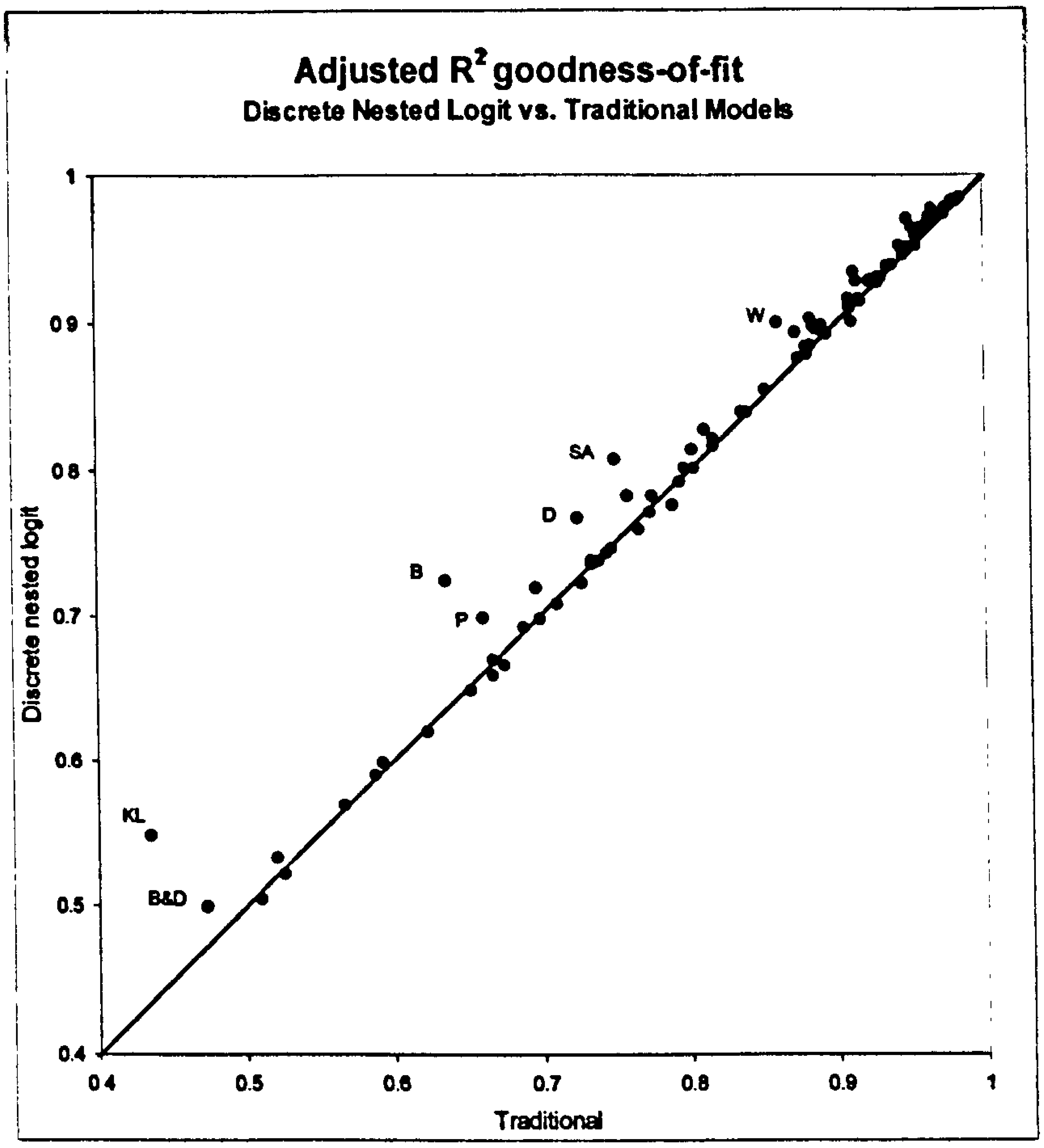


Figure 8.1: R^2_{adj} values from discrete nested logit and traditional models.

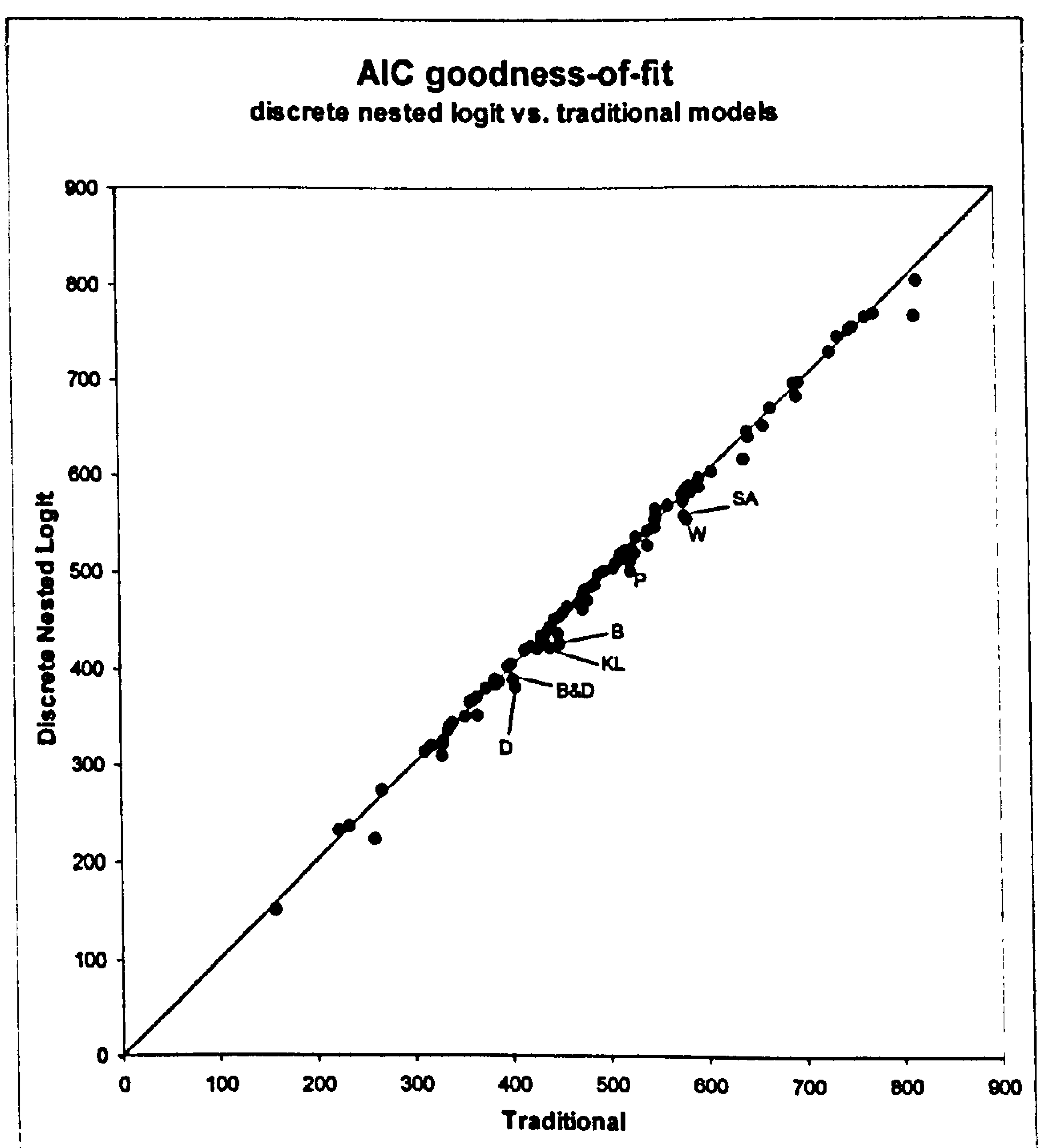


Figure 8.2: AIC values from discrete nested logit and traditional models.

Figure 8.2 confirms the general trend of goodness-of-fit improvements evident from figure 8.1. The distribution of goodness-of-fit change between traditional and discrete nested logit models is more clearly seen on a ranked bar chart, as in figure 8.3 below.

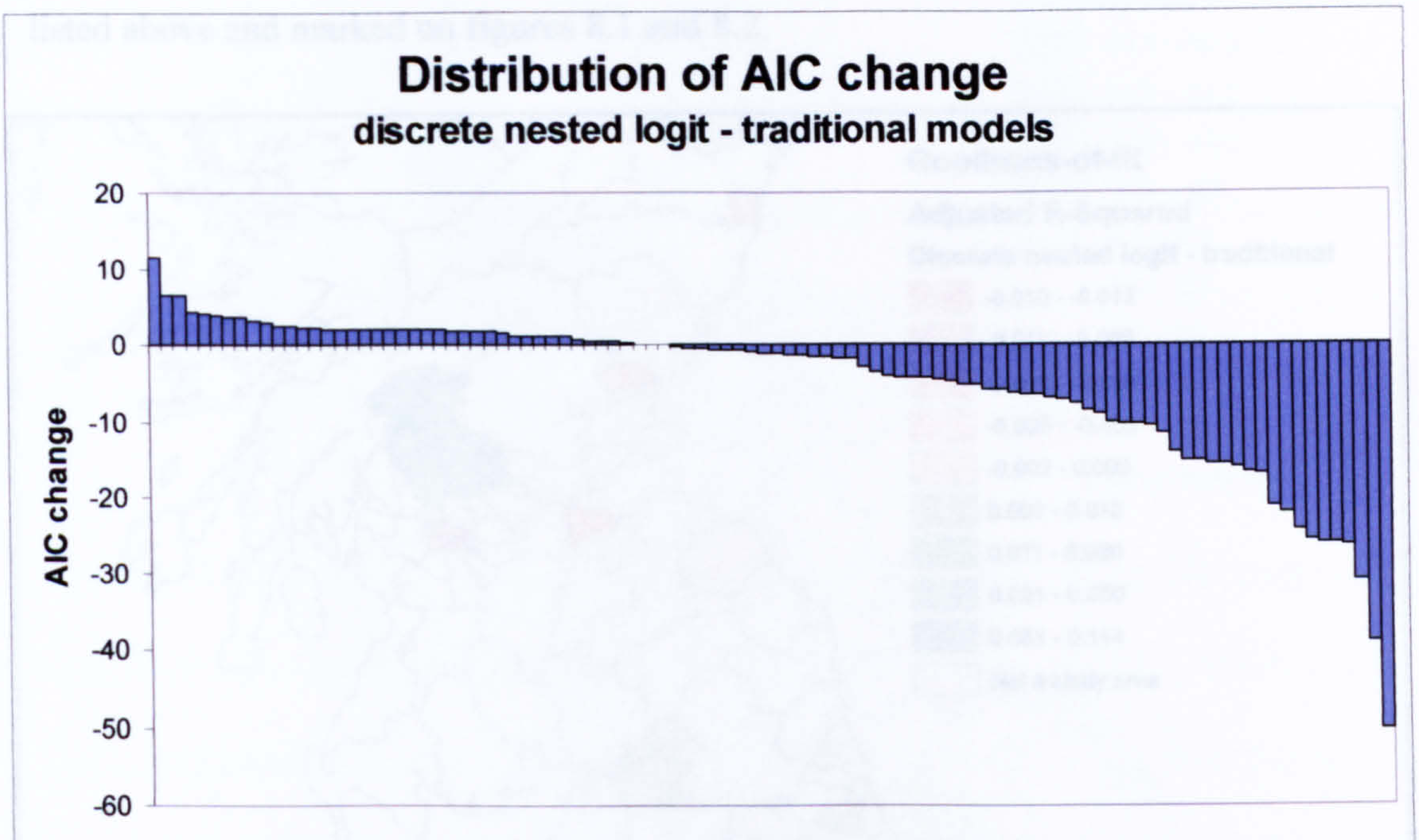
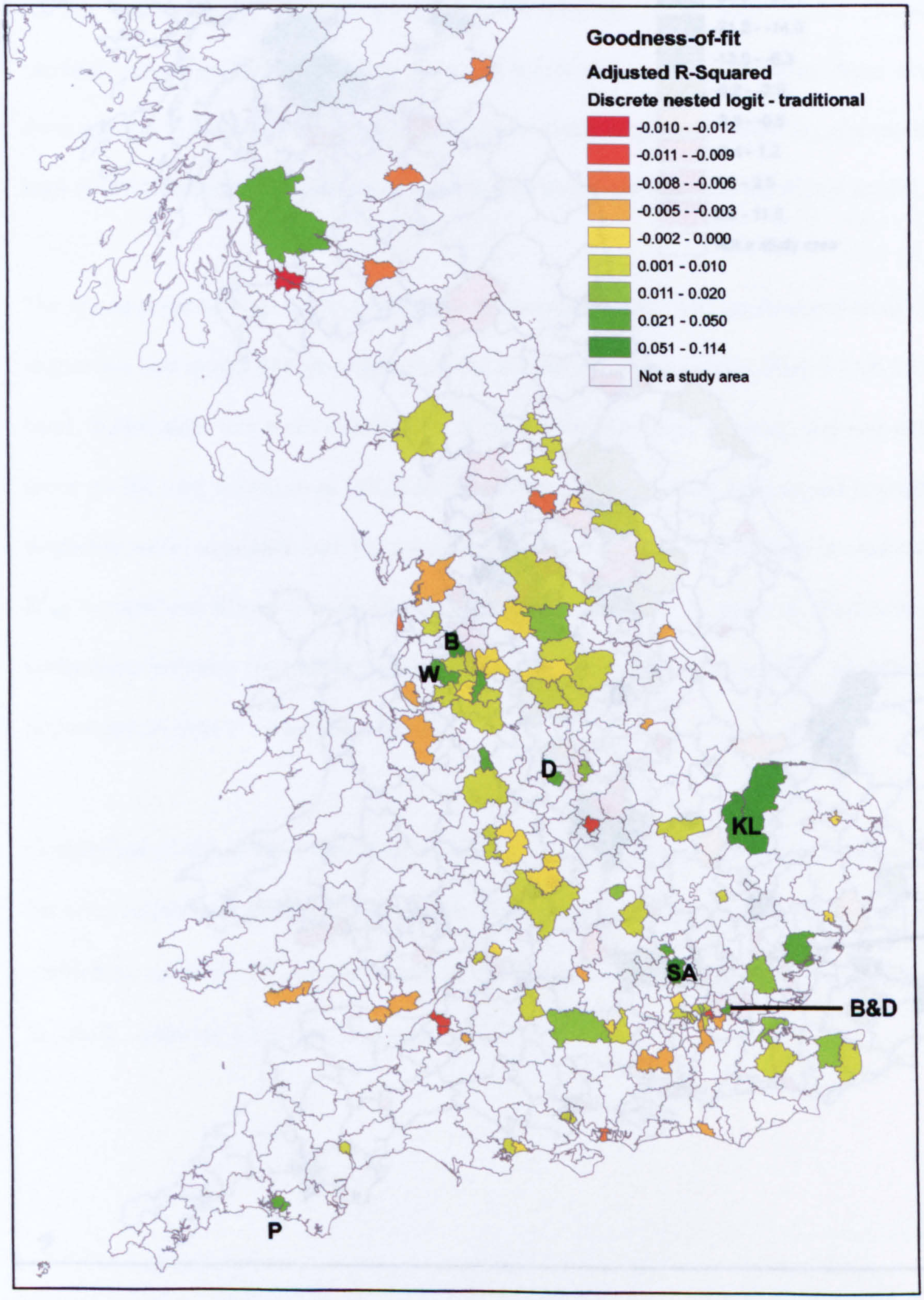


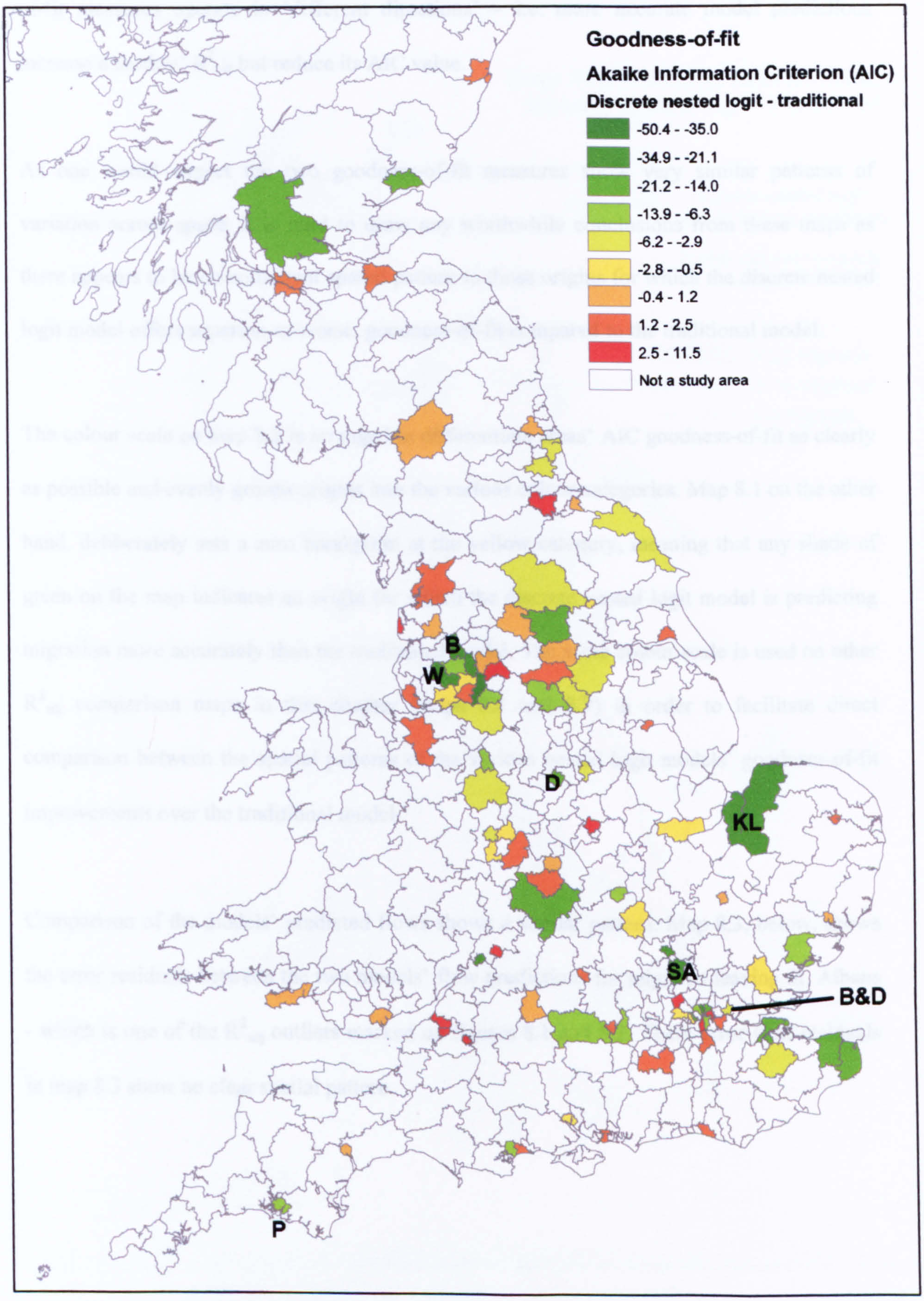
Figure 8.3: Distribution of percentage R^2_{adj} change, discrete nested logit – traditional models.

The most striking observation from figures 8.1 to 8.3 is that the improvements in goodness-of-fit of the discrete nested logit model are much more modest than those apparent from the competing destinations model, both in terms of the size of the improvements and the number of origins that experience an improvement in goodness-of-fit.

There appears to be no obvious spatial pattern to the origins for which the nested logit model provides improved predictive ability. The six origins whose discrete nested logit R^2_{adj} statistic is at least 5% higher than for the traditional model are: Kings Lynn & West Norfolk, Blackburn, St. Albans, Derby, Plymouth and Barking & Dagenham, (identified by their initials on figure 8.1 above) – these areas are labelled on both figures 8.1 and 8.2, above. These areas are geographically and socio-economically diverse, with no obvious commonalities to explain the superior predictive ability of the nested logit model for these origins. The possibility of a spatial pattern in the goodness-of-fit improvements is explored through maps 8.1 and 8.2, below, which plot R^2_{adj} and AIC change (discrete nested logit – traditional) for each of the 100 selected migration origins, and labels the seven outlier areas listed above and marked on figures 8.1 and 8.2.



Map 8.1: R^2_{adj} change, discrete nested logit – traditional model.



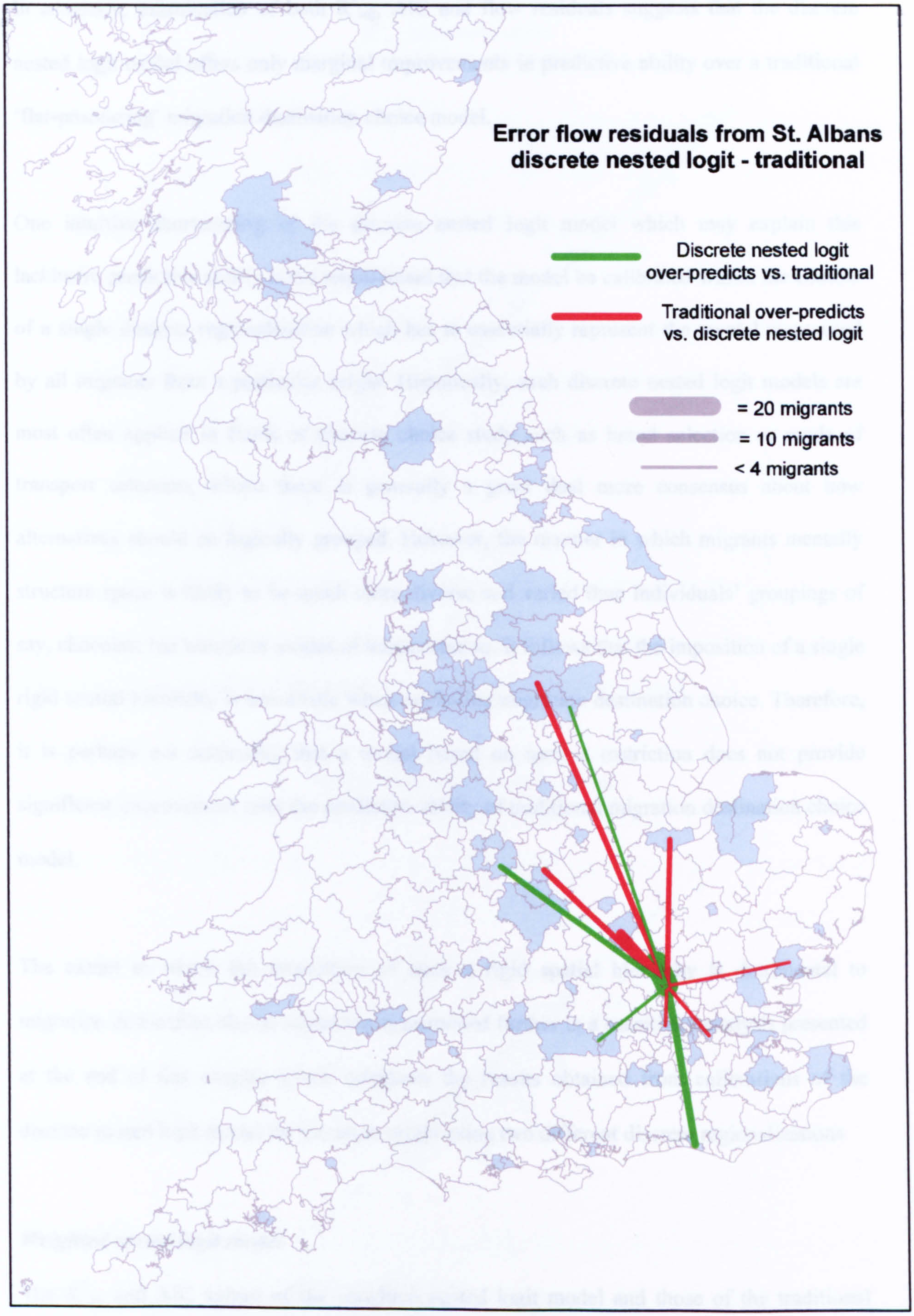
Map 8.2: AIC change, discrete nested logit – traditional model.

The colour scales on maps 8.1 and 8.2 are reversed to reflect the fact that the two goodness-of-fit measures operate in 'different directions' – i.e. more accurate model predictions increase a models' R^2_{adj} but reduce its AIC value.

As one would expect the two goodness-of-fit measures show very similar patterns of variation across space. It is hard to draw any worthwhile conclusions from these maps as there appears to be no consistent spatial pattern to those origins for which the discrete nested logit model offers superior, or worse, goodness-of-fit compared to the traditional model.

The colour scale on map 8.2 is arranged to differentiate areas' AIC goodness-of-fit as clearly as possible and evenly groups origins into the various colour categories. Map 8.1 on the other hand, deliberately sets a zero breakpoint at the yellow category, meaning that any shade of green on the map indicates an origin for which the discrete nested logit model is predicting migration more accurately than the traditional model. The same colour scale is used on other R^2_{adj} comparison maps in this chapter (maps 8.4 and 8.7) in order to facilitate direct comparison between the spatial patterns of the various nested logit models' goodness-of-fit improvements over the traditional model.

Comparison of the models' predicted flows shows a similar pattern. Map 8.3, below, shows the error residuals between the two models' flow predictions for migrants leaving St. Albans - which is one of the R^2_{adj} outliers marked on figures 8.1 and 8.2. These error flow residuals in map 8.3 show no clear spatial pattern.



Map 8.3: Residual flows from Derby: discrete nested logit vs. traditional.

In summary, examination of both R^2_{adj}, AIC and flow residuals suggests that the discrete nested logit model offers only marginal improvements in predictive ability over a traditional 'flat-processing' migration destination choice model.

One intuitive shortcoming of the discrete nested logit model which may explain this lacklustre predictive ability is the requirement that the model be calibrated within the context of a single discrete regionalization which has to essentially represent the mental maps used by all migrants from a particular origin. Historically, such discrete nested logit models are most often applied in fields of discrete choice study such as brand selection or mode of transport selection, where there is generally a great deal more consensus about how alternatives should be logically grouped. However, the manner in which migrants mentally structure space is likely to be much more diverse and varied than individuals' groupings of say, chocolate bar brands or modes of transportation. It follows that the imposition of a single rigid spatial hierarchy is unrealistic when modelling migration destination choice. Therefore, it is perhaps not surprising that a model based on such a restriction does not provide significant improvement over the predictive ability of traditional migration destination choice model.

The extent to which the imposition of such a rigid spatial hierarchy is detrimental to migration destination choice modelling is examined further in a sensitivity analysis presented at the end of this chapter which compares the results obtained from calibrations of the discrete nested logit model for the same origin using two different discrete regionalizations.

Weighted nested logit model

The R^2_{adj} and AIC values of the weighted nested logit model and those of the traditional model are compared in figure 8.4 and 8.5 below:

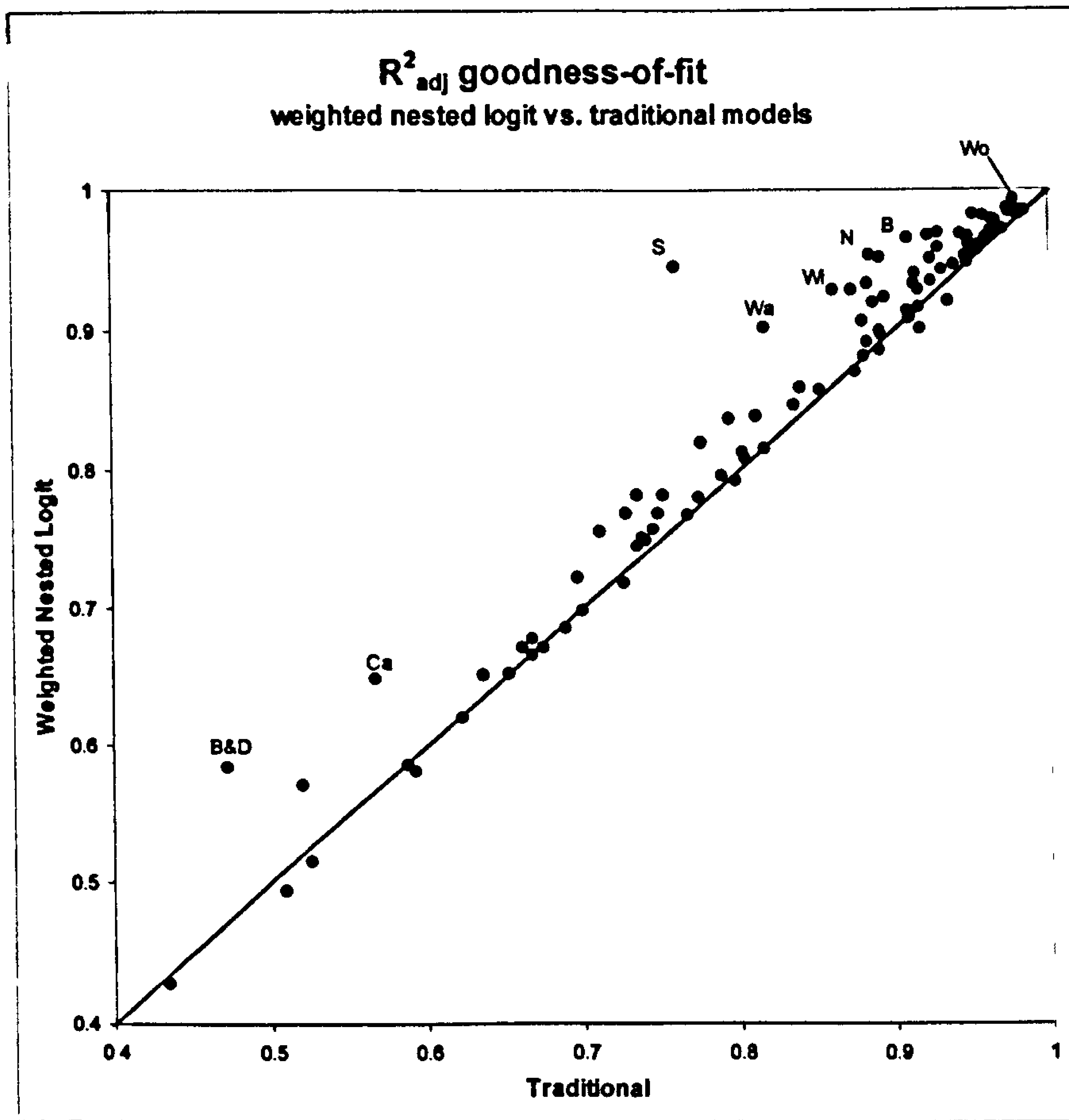


Figure 8.4: R^2_{adj} statistics for weighted nested logit and traditional models.

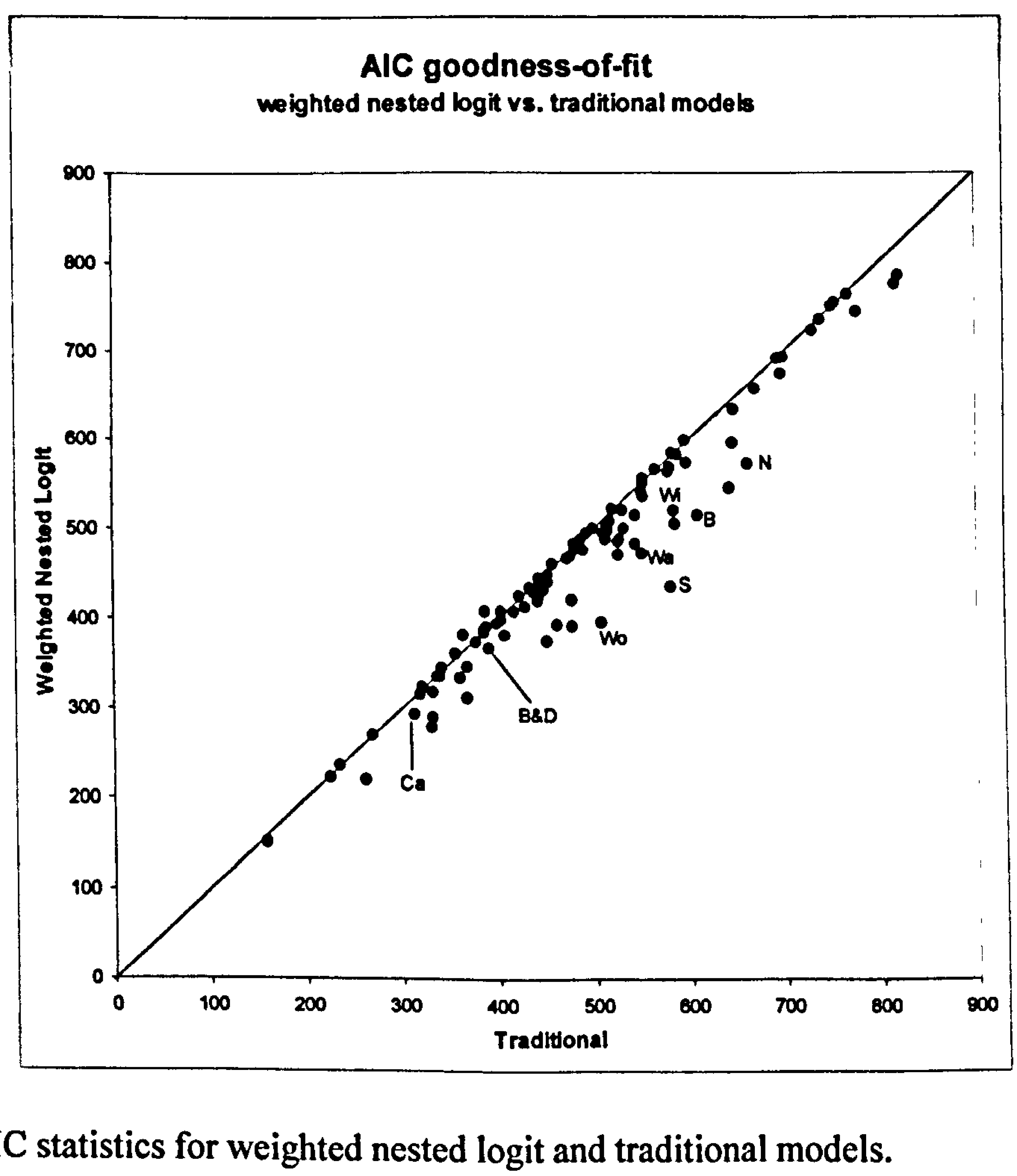


Figure 8.5: AIC statistics for weighted nested logit and traditional models.

If figures 8.4 and 8.5 are compared with figures 8.1 and 8.2 it is immediately evident that there is more variation in the R^2_{adj} and AIC values between the weighted nested logit and traditional models, than was shown above between the discrete nested logit and traditional models. There are many more origins for which the weighted nested logit model out-performs both the traditional model and the discrete nested logit model.

The areas showing weighted nested logit R^2_{adj} values which are significantly (>10%) higher than the traditional model are: Greenwich, Kings Lynn & West Norfolk, Stoke-on-Trent, St. Albans, Barking & Dagenham, Chelmsford and Northampton. These areas are marked on figures 8.4 and 8.5, and once again can generally be seen to also be amongst the better performing areas as indicated by improvement (i.e. reduction) in their AIC values.

These outlier origins for the weighted nested logit model show little correlation with the best improvers from the competing destinations model (see figure 7.2). Whilst the competing destinations model's best improver, Kings Lynn & West Norfolk, is also a nested logit outlier, other top nested logit performers, such as Barking & Dagenham and Stoke-on-Trent, have competing destinations R^2_{adj} values that are virtually unchanged from their traditional model R^2_{adj} values. It should be noted that total out-migration from Kings Lynn and West Norfolk is one of the lowest of the group of 100 origins, so it is particularly impressive to obtain such good model fit given the limited data against which the model is being calibrated. All other weighted nested logit 'improvers' have reasonable out-migration, particularly Greenwich which showed the most improvement.

Of the top seven weighted nested logit model performers, only three are in common with those from the discrete nested logit. This suggests that the probabilistic regionalization used when calibrating the weighted nested logit model is capturing migrants' mental maps of space in a different manner to the discrete nested logit model's discrete regionalizations.

Furthermore, the two approaches appear to be appropriate to migrants from different origins, though it is likely that the best discrete nested logit performers would vary if the model was recalibrated against another set of discrete regionalizations.

The statistical distribution of values of change in AIC between the traditional model and the weighted nested logit model is presented in figure 8.6 below.

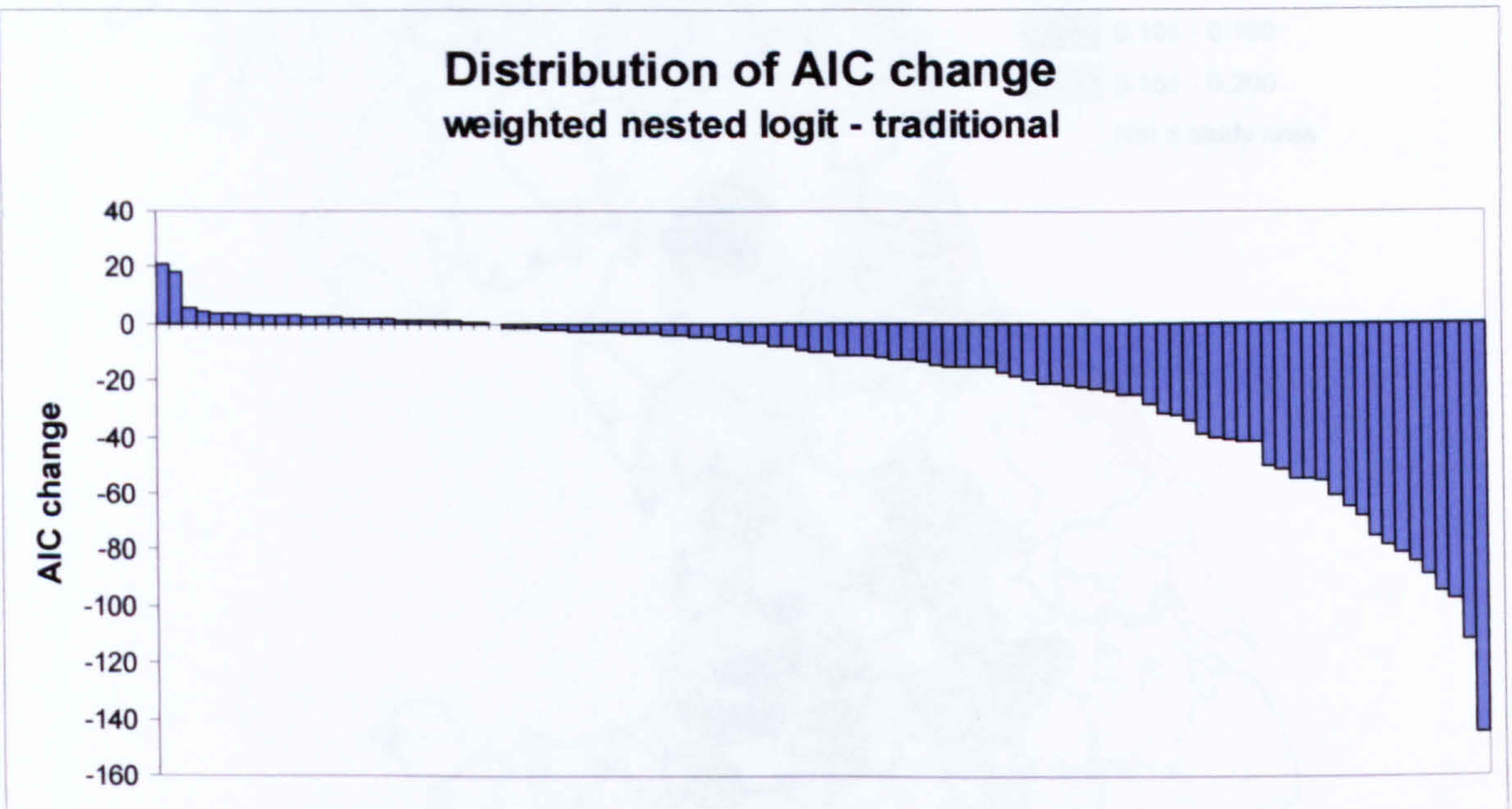
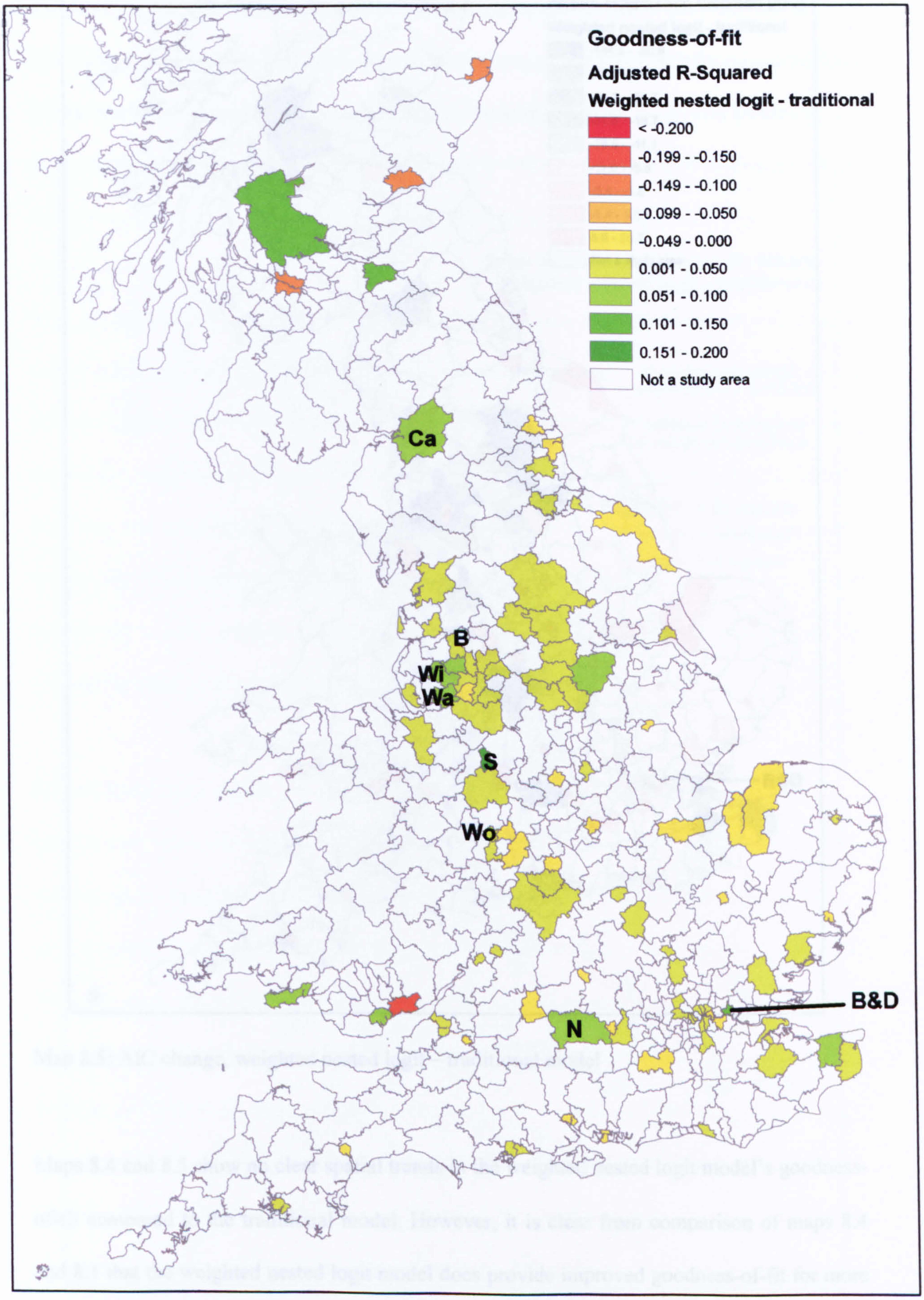


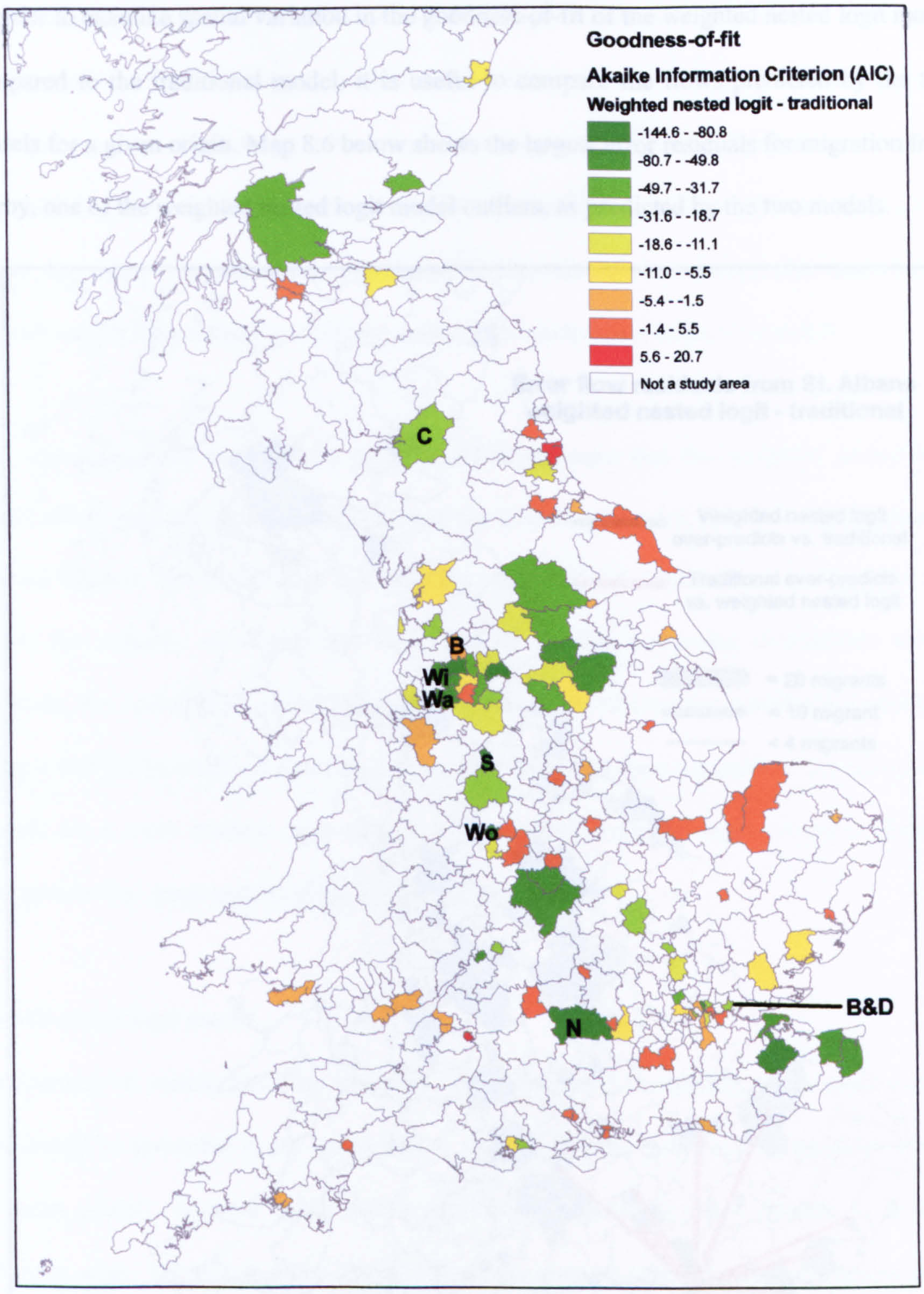
Figure 8.6: Distribution of change in AIC, weighted nested logit – traditional models.

When figure 8.6 is compared with figure 8.3 above, it is evident that the weighted nested logit model produces improvements in goodness-of-fit when modelling migration from more of the 100 origins, and also that the amount of improvement is more substantial for the weighted discrete nested logit model.

The spatial distribution of change in goodness-of-fit between the weighted nested logit model and the traditional model are examined in maps 8.4 and 8.5 below, which plot change in R^2_{adj} and AIC, respectively, and label the seven outliers identified on figures 8.4 and 8.5.



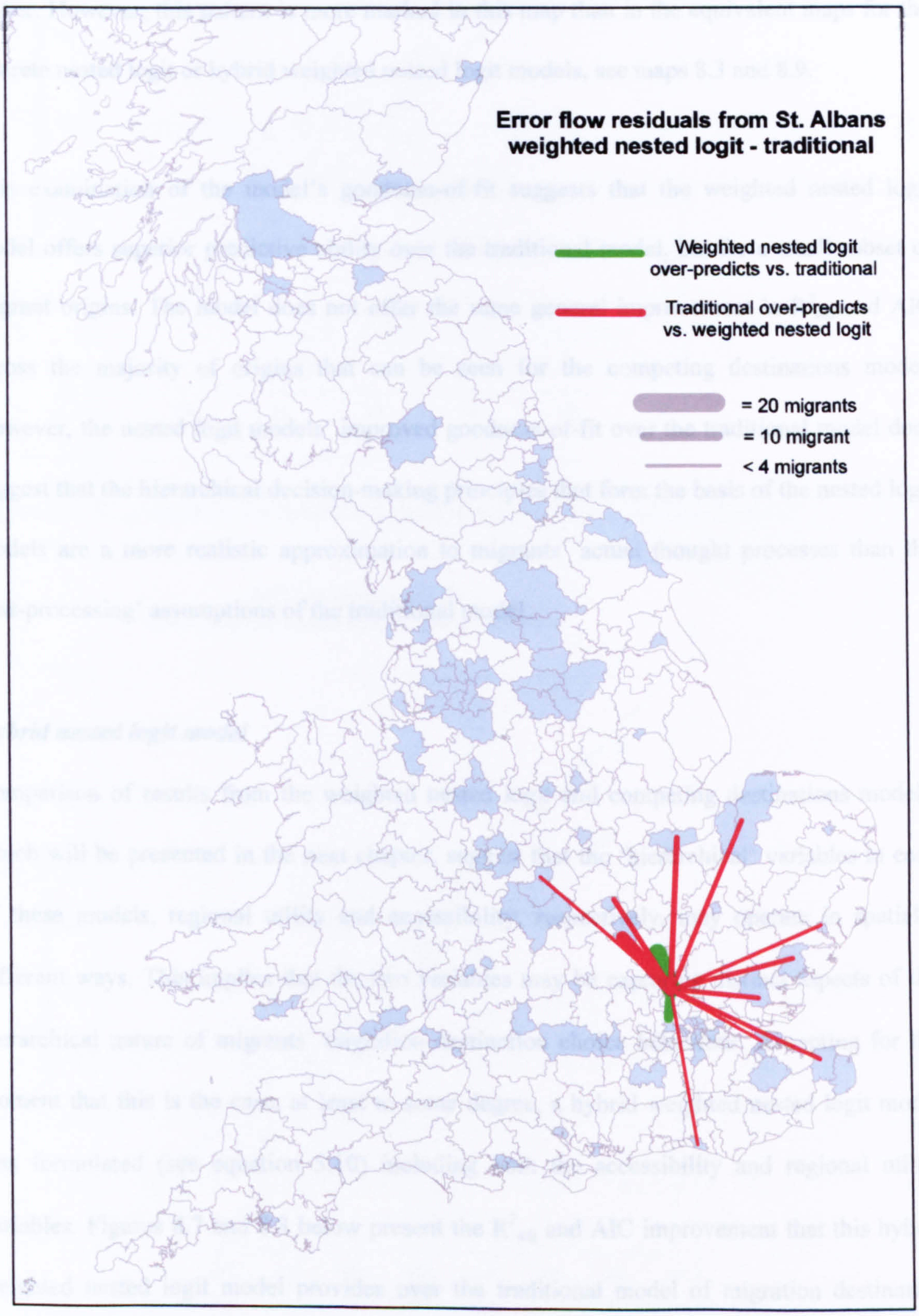
Map 8.4: R^2_{adj} change, weighted nested logit – traditional model



Map 8.5: AIC change, weighted nested logit – traditional model

Maps 8.4 and 8.5 show no clear spatial trends in the weighted nested logit model's goodness-of-fit compared to the traditional model. However, it is clear from comparison of maps 8.4 and 8.1 that the weighted nested logit model does provide improved goodness-of-fit for more migration origins throughout the country, compared to the discrete nested logit model.

In order to examine spatial variation in the goodness-of-fit of the weighted nested logit model compared to the traditional model, it is useful to compare the flows predicted by the two models for a given origin. Map 8.6 below shows the largest error residuals for migration from Derby, one of the weighted nested logit model outliers, as predicted by the two models.



Map 8.6: Residual flows from Derby: weighted nested logit vs. traditional models.

It is interesting to note from map 8.6 that all of the top 13 error flow residuals are quite short-distance moves. This is perhaps not entirely surprising given that we are considering actual error (as a count of migrants) rather than percentage error, and flows to closer areas are generally larger, so the errors in flows to closer destinations might be expected to be a little higher. However, this pattern is more marked in this map than in the equivalent maps for the discrete nested logit or hybrid weighted nested logit models, see maps 8.3 and 8.9.

This examination of the model's goodness-of-fit suggests that the weighted nested logit model offers superior predictive ability over the traditional model, but for a small subset of migrant origins. The model does not offer the same general improvement in R^2_{adj} and AIC across the majority of origins that can be seen for the competing destinations model. However, the nested logit models' improved goodness-of-fit over the traditional model does suggest that the hierarchical decision-making principles that form the basis of the nested logit models are a more realistic approximation to migrants' actual thought processes than the 'flat-processing' assumptions of the traditional model.

Hybrid nested logit model

Comparison of results from the weighted nested logit and competing destinations models, which will be presented in the next chapter, suggest that the 'hierarchical' variables in each of these models, regional utility and accessibility, respectively, may operate in spatially different ways. This implies that the two variables may be capturing distinct aspects of the hierarchical nature of migrants' migration destination choice behaviour. Accepting for the moment that this is the case, at least to some degree, a hybrid weighted nested logit model was formulated (see equation 5.10) including both the accessibility and regional utility variables. Figures 8.7 and 8.8 below present the R^2_{adj} and AIC improvement that this hybrid weighted nested logit model provides over the traditional model of migration destination choice.

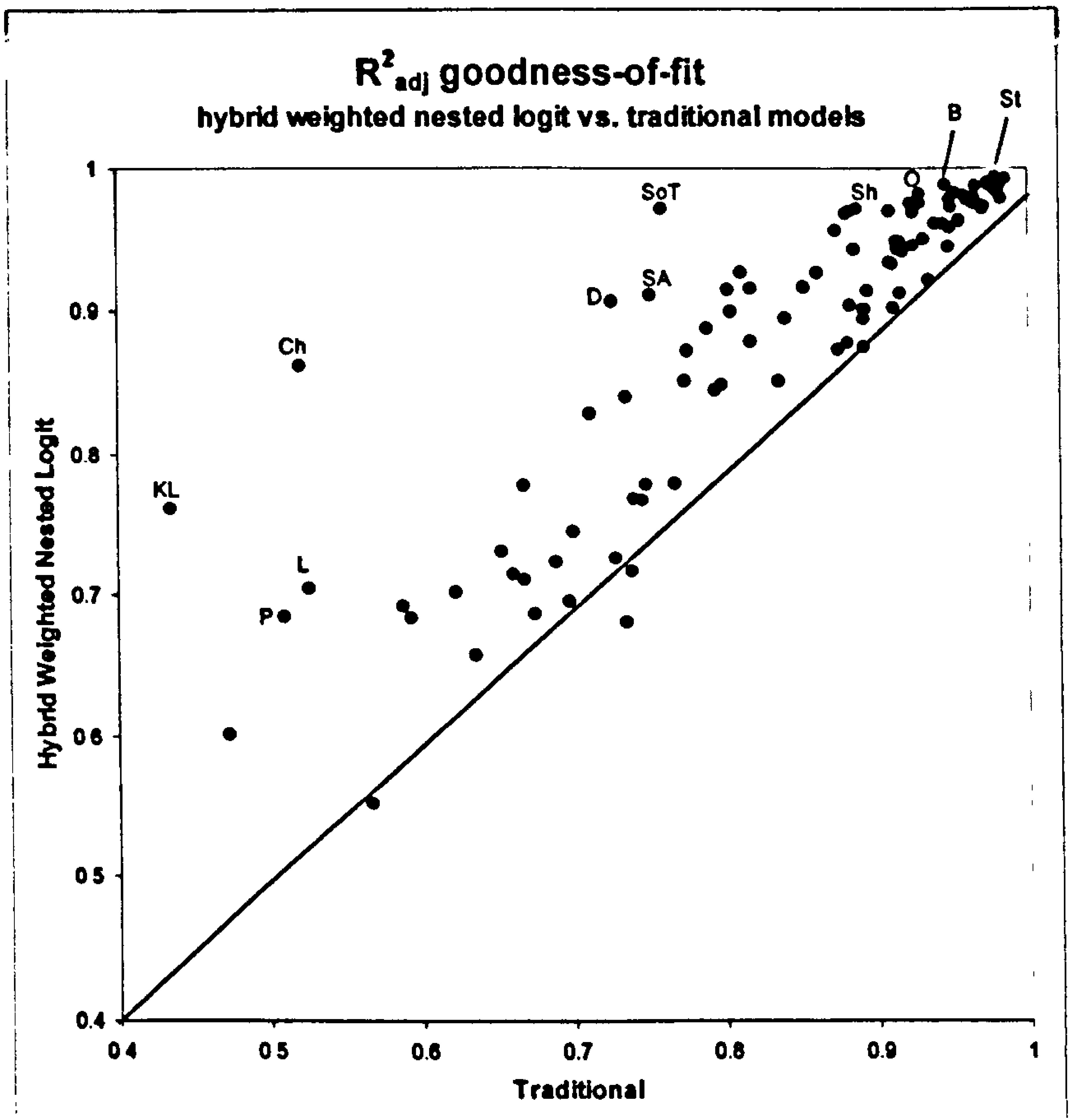


Figure 8.7: R^2_{adj} statistics for hybrid weighted nested logit and traditional models.

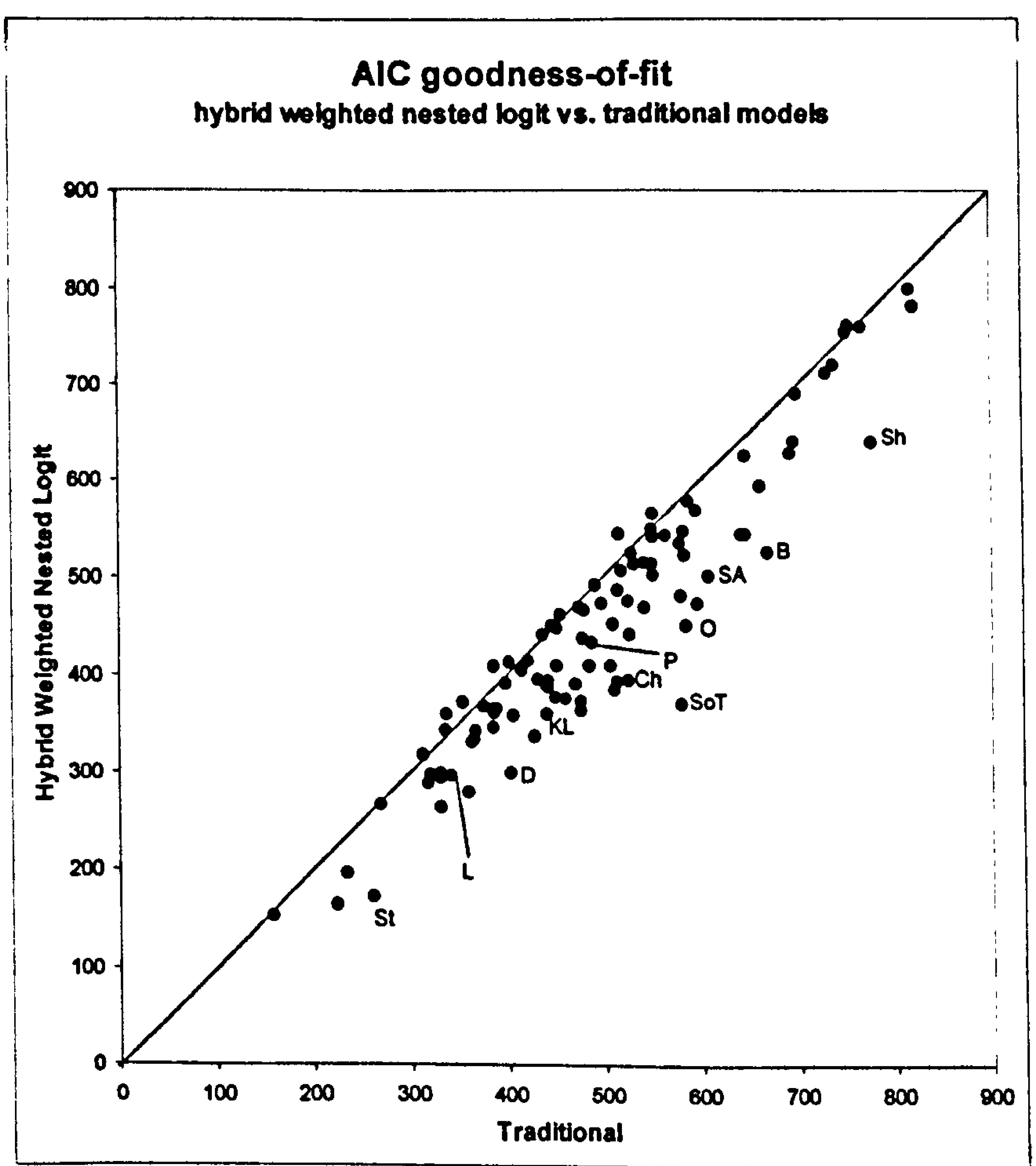


Figure 8.8: AIC statistics for hybrid weighted nested logit and traditional models.

It is evident from figures 8.7 and 8.8 that, of all the models calibrated here, the hybrid weighted nested logit model provides the best goodness-of-fit. This supports the assumption, made above, that the accessibility and regional utility variables are capturing different aspects of migrants' hierarchical decision-making, and justifies the formulation of a new hybrid destination choice model combining the benefits of both variables.

The almost universal improvement in goodness-of-fit that the hybrid weighted nested logit model offers over the traditional model can be effectively visualized through the statistical distribution of the changes in AIC between the models – as presented in figure 8.9, below.

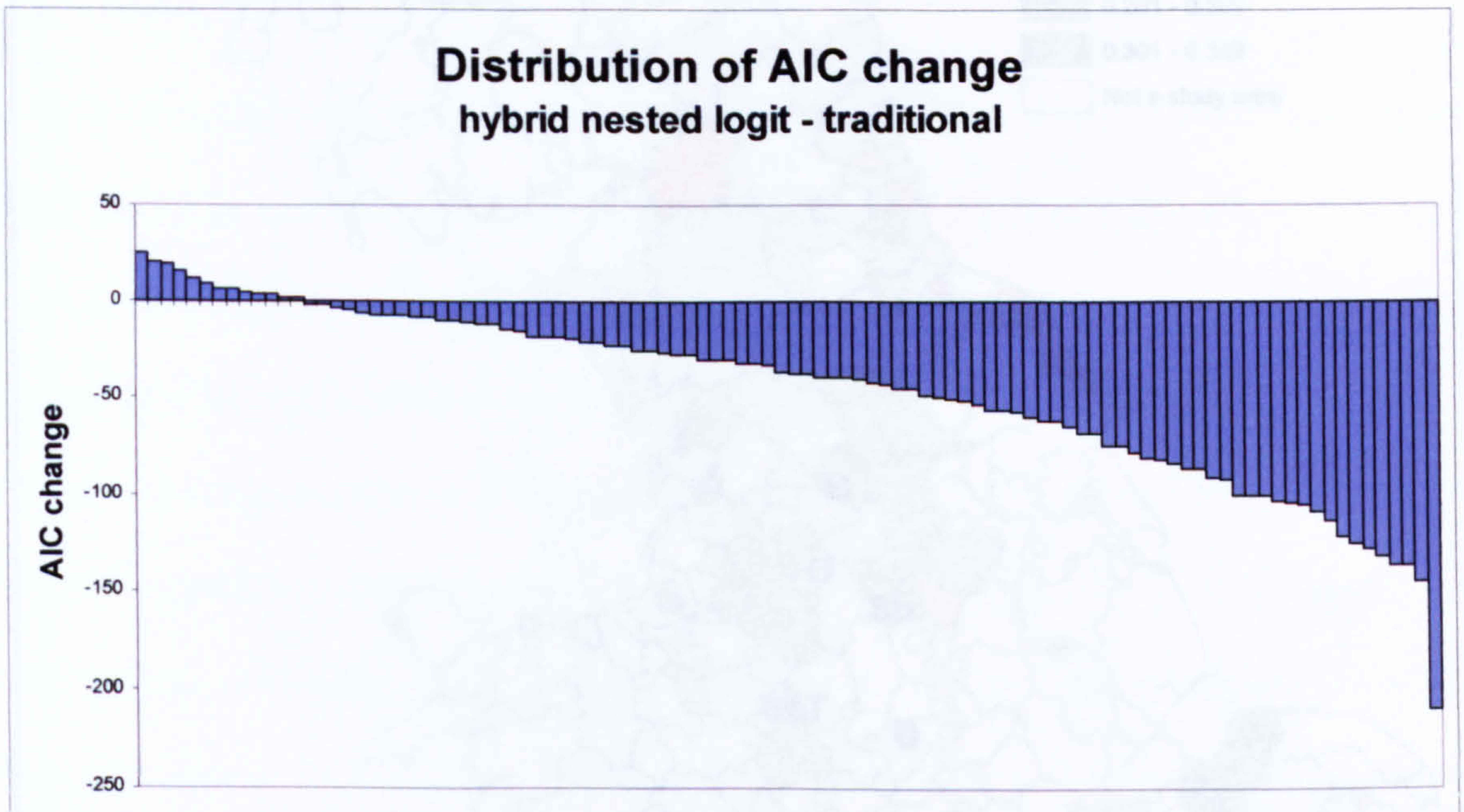
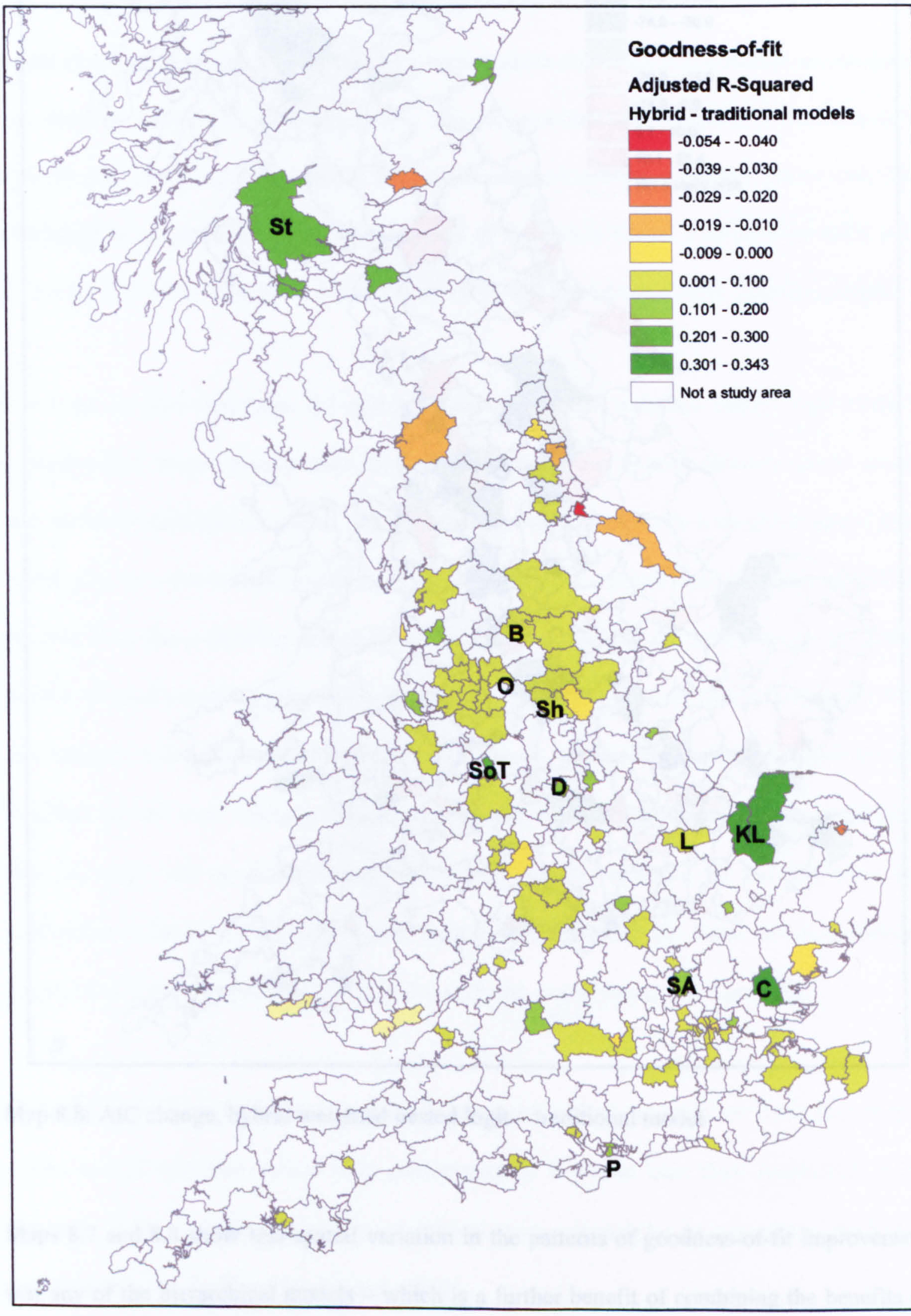


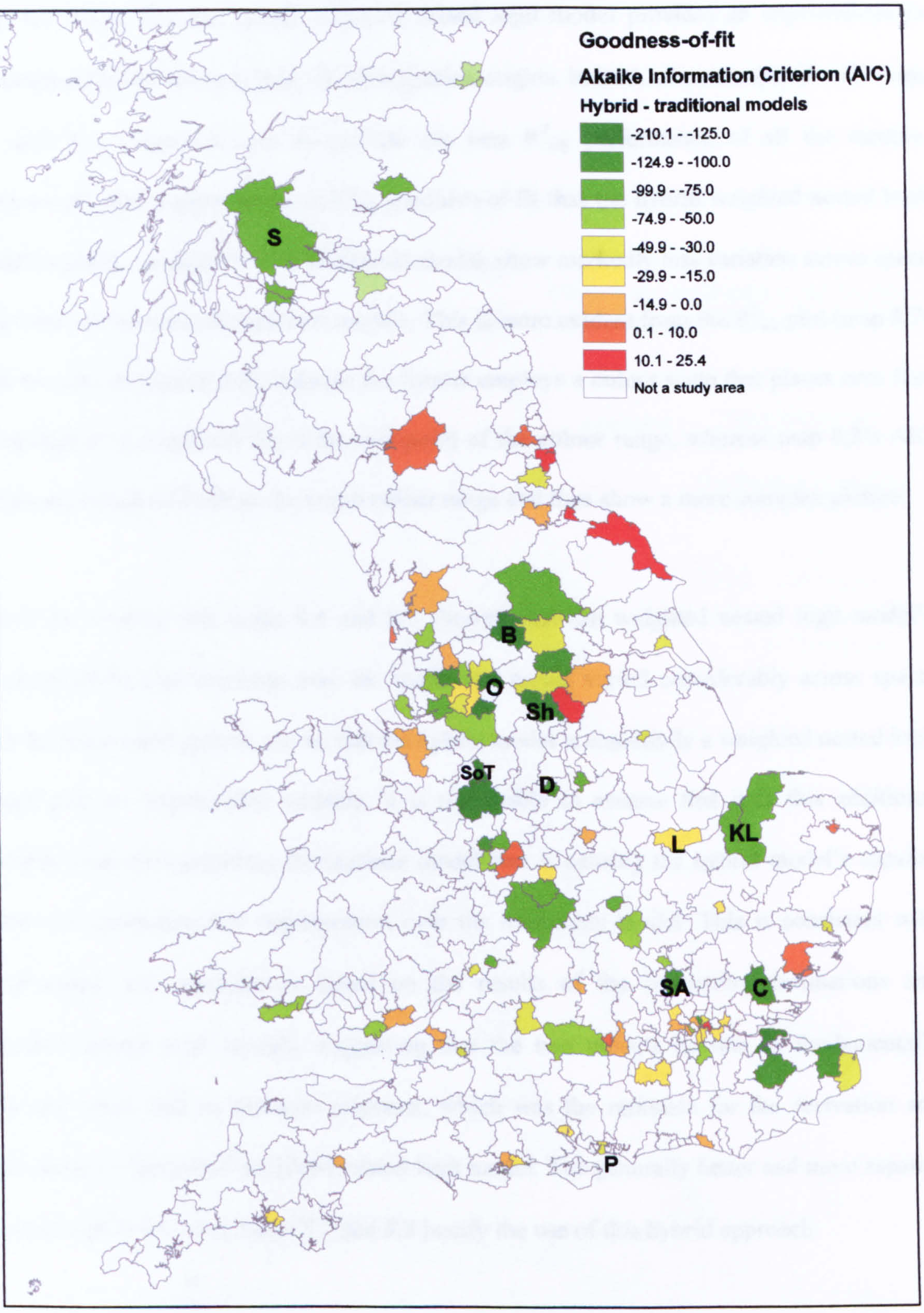
Figure 8.9: Distribution of changes in AIC, hybrid weighted nested logit–traditional models.

Comparison of figure 8.9 with figures 8.3 and 8.6 shows the superior goodness-of-fit of the hybrid model provides supporting the proposition above that the accessibility variable from the competing destinations model and the regional utility variable from the weighted nested logit model capture the hierarchical aspects of migrants' destination choice processes in inherently different ways, and justifying the combination of these two variables together in the hybrid model. This will be discussed further in chapter 9 which presents a variety of comparisons between the hierarchical models introduced and applied in this research.

Again, spatial patterns in the hybrid model's goodness-of-fit improvements over the traditional model are explored by mapping the differences in R^2_{adj} and AIC statistics between these models for each of the 100 selected origins, see maps 8.7 and 8.8.



Map 8.7: R^2_{adj} change, hybrid weighted nested logit – traditional model.



Map 8.8: AIC change, hybrid weighted nested logit – traditional model

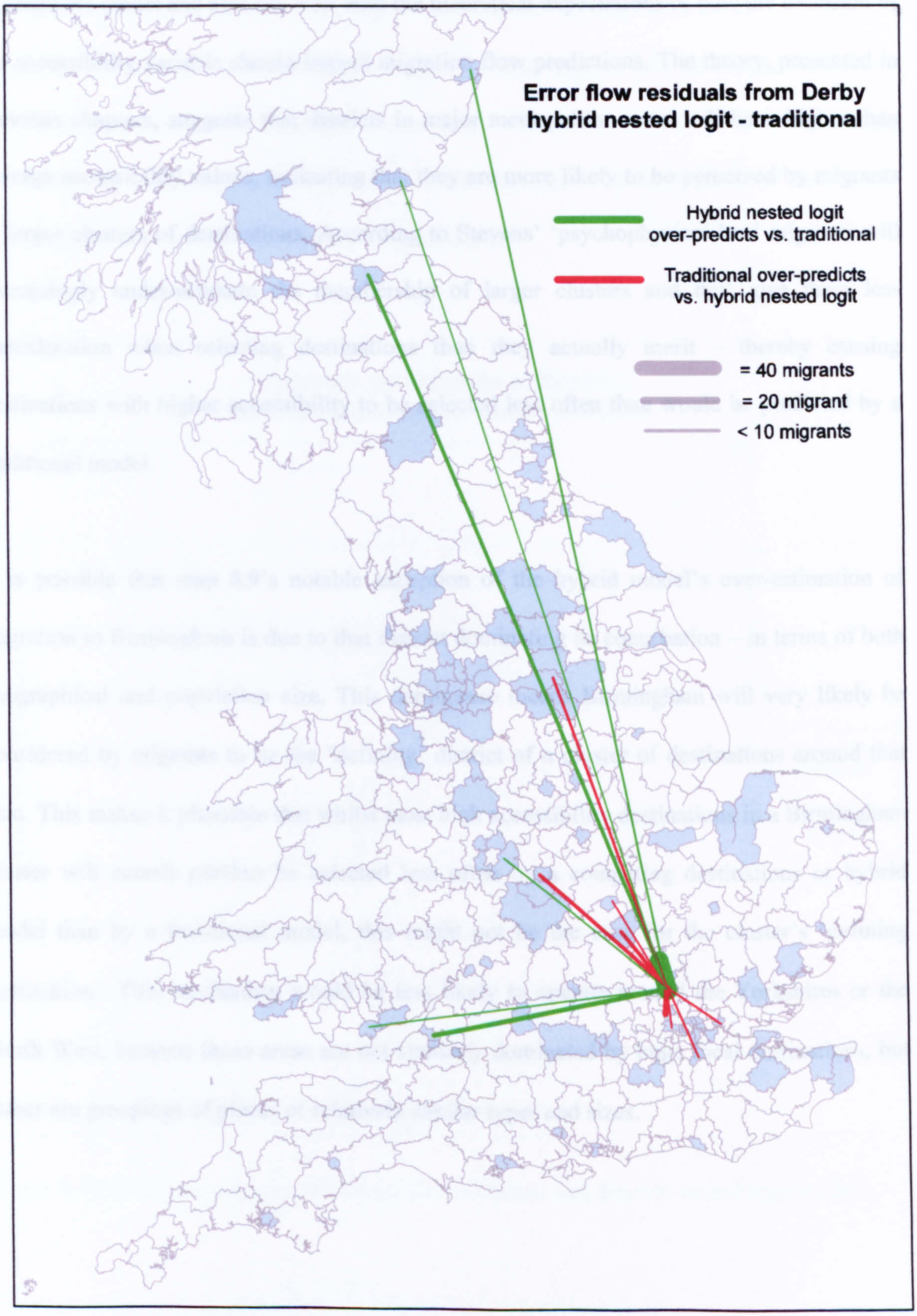
Maps 8.7 and 8.8 show less spatial variation in the patterns of goodness-of-fit improvement that any of the hierarchical models – which is a further benefit of combining the benefits of the competing destinations and weighted nested logit models.

Map 8.7 shows that the hybrid weighted nested logit model provides an improvement in goodness-of-fit for the vast majority of migration origins. Indeed, when compared with maps 8.1 and 8.4, it can be seen to provide the best R^2_{adj} performance of all the models. Furthermore, the improvements in R^2_{adj} goodness-of-fit that the hybrid weighted nested logit model exhibits, compared to the traditional model, show markedly less variation across space than those of the other hierarchical models. This is more evident from the R^2_{adj} plot (map 8.7) than the AIC plot (map 8.8), because the former employs a colour scale that places zero (i.e. no change in goodness-of-fit) at the mid-point of the colour range, whereas map 8.8's AIC values are distributed across the entire colour range and thus show a more complex picture.

It will be recalled that maps 8.4 and 8.5 showed that the weighted nested logit model's goodness-of-fit improvements over the traditional model varied considerably across space, with no discernable pattern. Given that the hybrid model is essentially a weighted nested logit model plus an accessibility variable, it is reasonable to assume that it is this additional variable from the competing destinations model that is causing the hybrid model's aspatial pattern of goodness-of-fit improvement over the traditional model. This is consistent with observations and speculations based on the results of the competing destinations and weighted nested logit models, suggesting that the two models operate in fundamentally different ways, and on different migrants, which was the rationale for the derivation and application of the hybrid weighted nested logit model. The generally better and more aspatial goodness-of-fit shown in maps 8.7 and 8.8 justify the use of this hybrid approach.

Maps 8.7 and 8.8 also suggest that a slight tendency appears to exist for those few origins whose hybrid weighted nested logit goodness-of-fit is worse than their traditional model goodness-of-fit, to be located in more peripheral areas – though by no means does the hybrid model perform badly for all peripheral origins.

It is also interesting to examine the error flow residuals, for a particular migration origin, between the predictions of the traditional model and those from the novel hybrid weighted nested logit model – these are presented for the origin Derby in map 8.9 below.



Map 8.9: Error flow residuals for migration from Derby, hybrid - traditional models.

The clear trend evident from map 8.9 is that the hybrid model tends to over-predict migration to more peripheral areas and generally under-predicts migration to destinations in more metropolitan areas. This is consistent with the behaviour observed for the competing destinations model and also fits well with the theoretical expectations of how the inclusion of the accessibility variable should impact migration flow predictions. The theory, presented in previous chapters, suggests that districts in major metropolitan areas will have higher than average accessibility values, indicating that they are more likely to be perceived by migrants in larger clusters of destinations. According to Stevens' 'psychophysical law' migrants will increasingly under-estimate the membership of larger clusters and thus give them less consideration when selecting destinations than they actually merit – thereby causing destinations with higher accessibility to be selected less often than would be predicted by a traditional model.

It is possible that map 8.9's notable exception of the hybrid model's over-estimation of migration to Birmingham is due to that district dominating its conurbation – in terms of both geographical and population size. This dominance means Birmingham will very likely be considered by migrants to be the 'defining' district of a cluster of destinations around that area. This makes it plausible that whilst most high accessibility destinations in a Birmingham cluster will *ceteris paribus* be selected less often by a competing destinations or hybrid model than by a traditional model, this might not be the case for the cluster's 'defining destination'. This mechanism would be less likely to operate around the Yorkshires or the North West, because those areas are not similarly dominated by huge focal destinations, but rather are groupings of places of relatively similar types and sizes.

Examination of Parameter Estimates

Discrete nested logit model

The parameter estimates generated by the discrete nested logit model show a very strong correlation with those generated by the traditional model. Figures 8.10 to 8.15 plot the parameter estimates of the traditional and discrete nested logit models for the six explanatory variables. A summary of the parameter estimates' correlation coefficients is listed in table 8.1. The parameter estimates (and goodness-of-fit statistics) from calibrations of the discrete nested logit model are tabulated in full in appendix F.

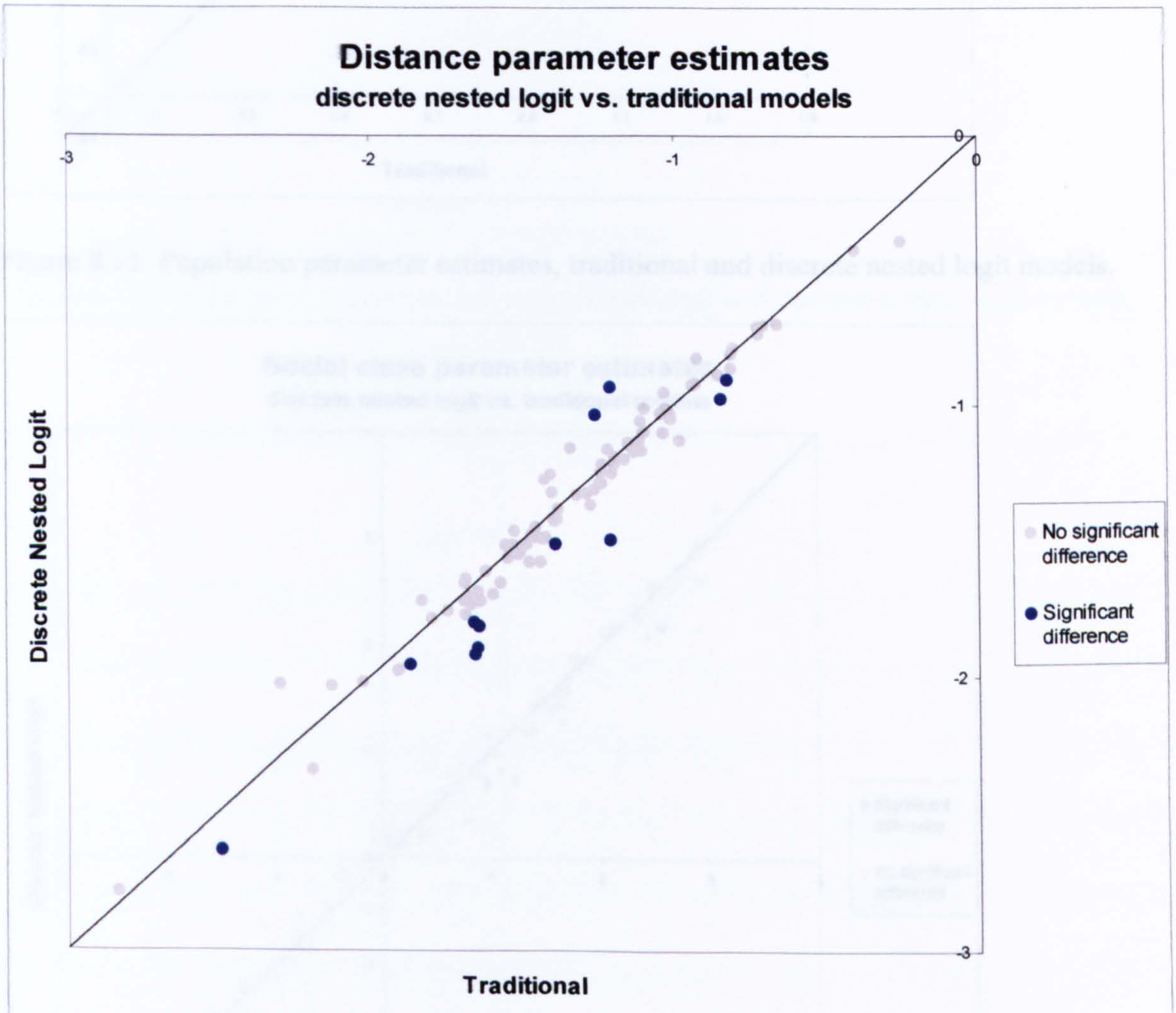


Figure 8.10: Distance parameter estimates for traditional and discrete nested logit models.

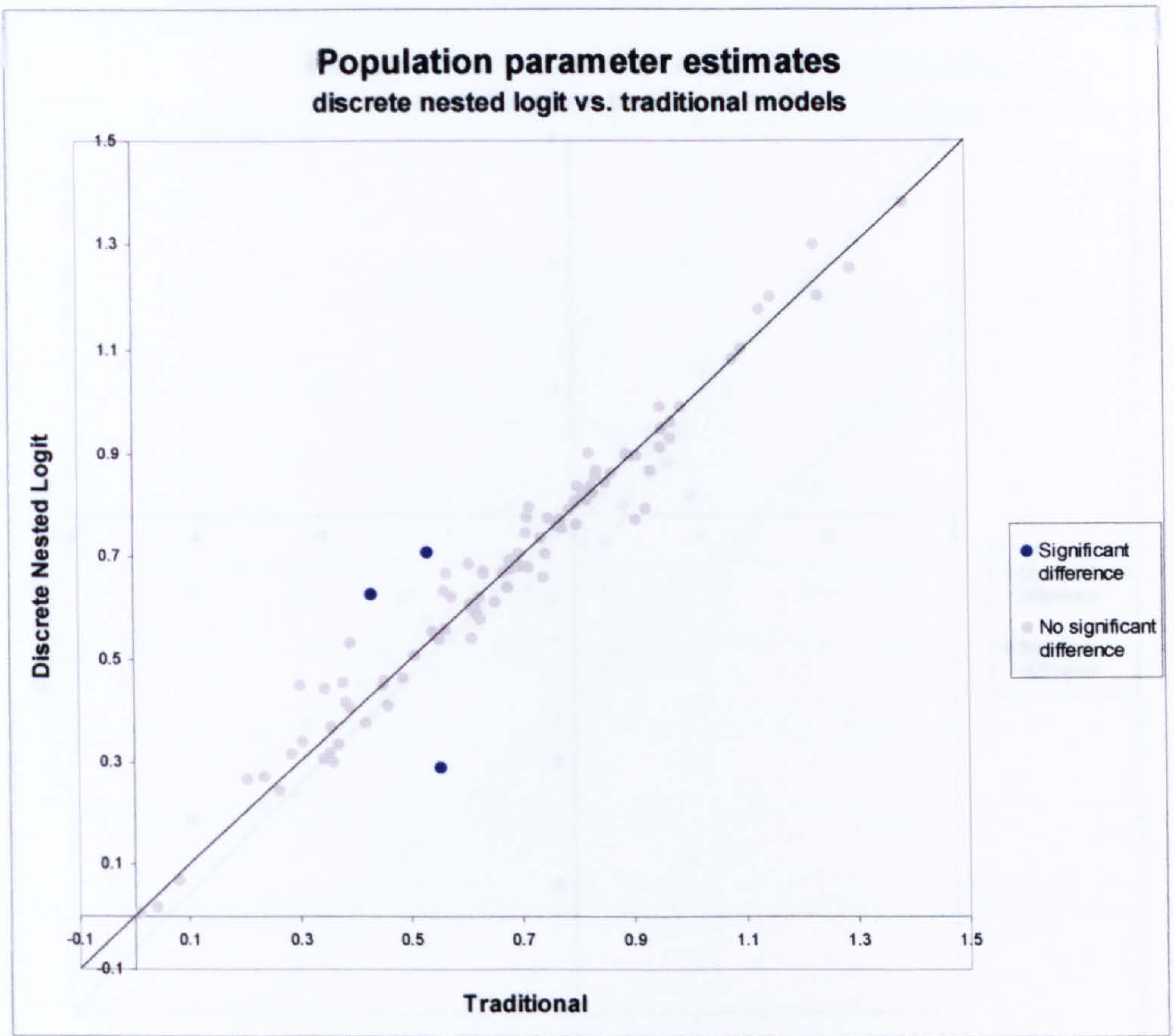


Figure 8.11: Population parameter estimates, traditional and discrete nested logit models.

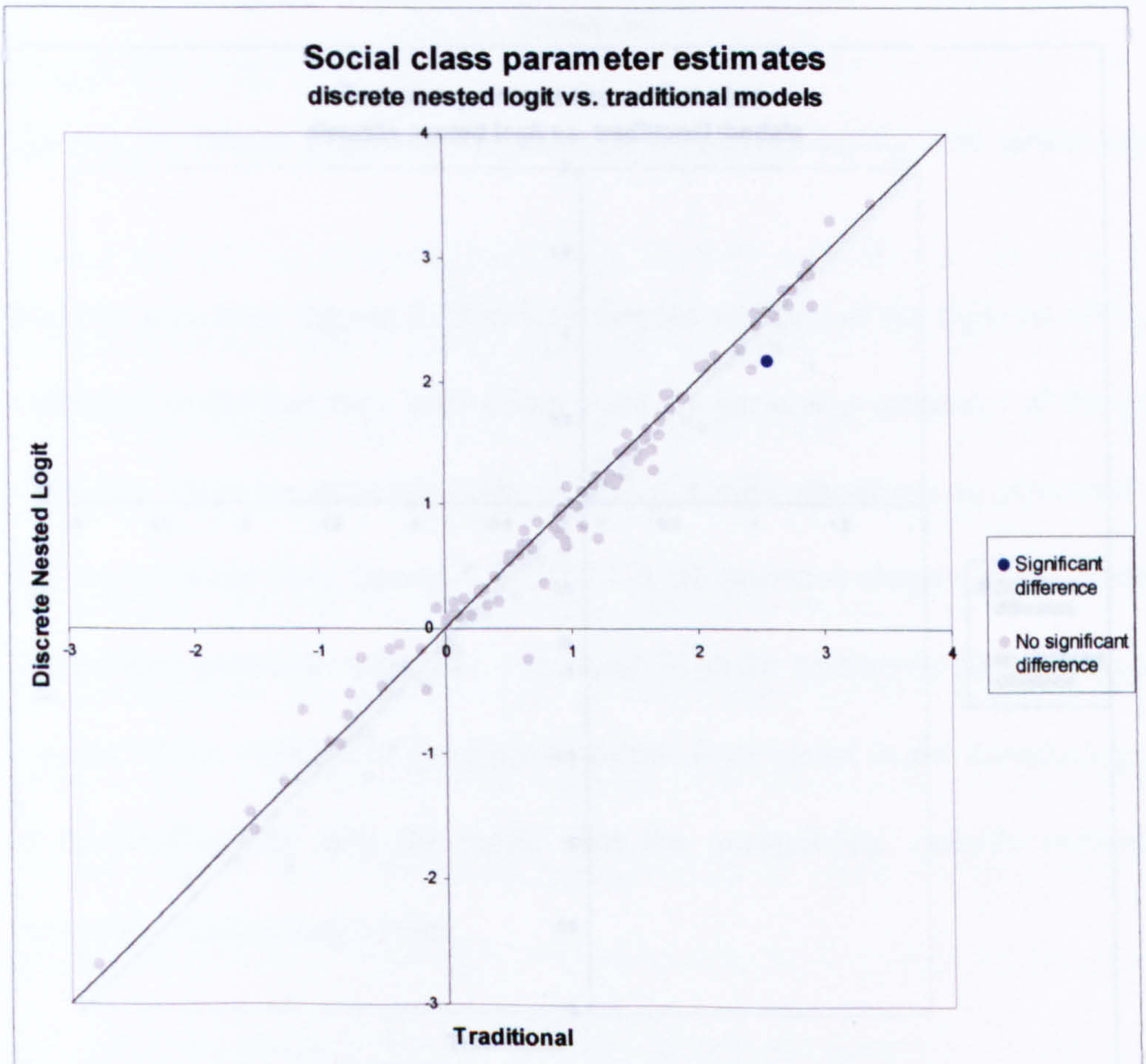


Figure 8.12: Social class parameter estimates, traditional and discrete nested logit models.

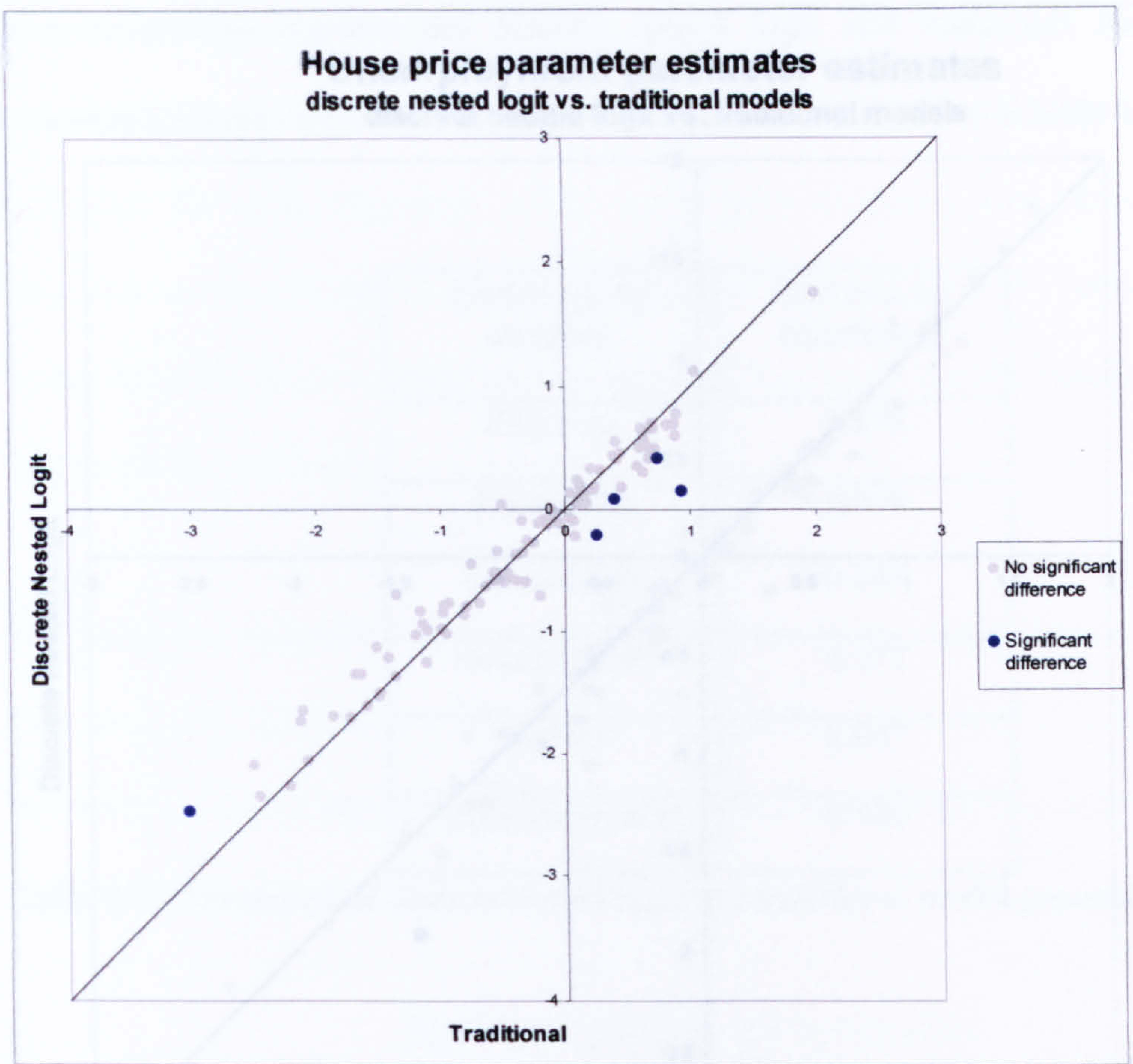


Figure 8.13: House price parameter estimates, traditional and discrete nested logit models.

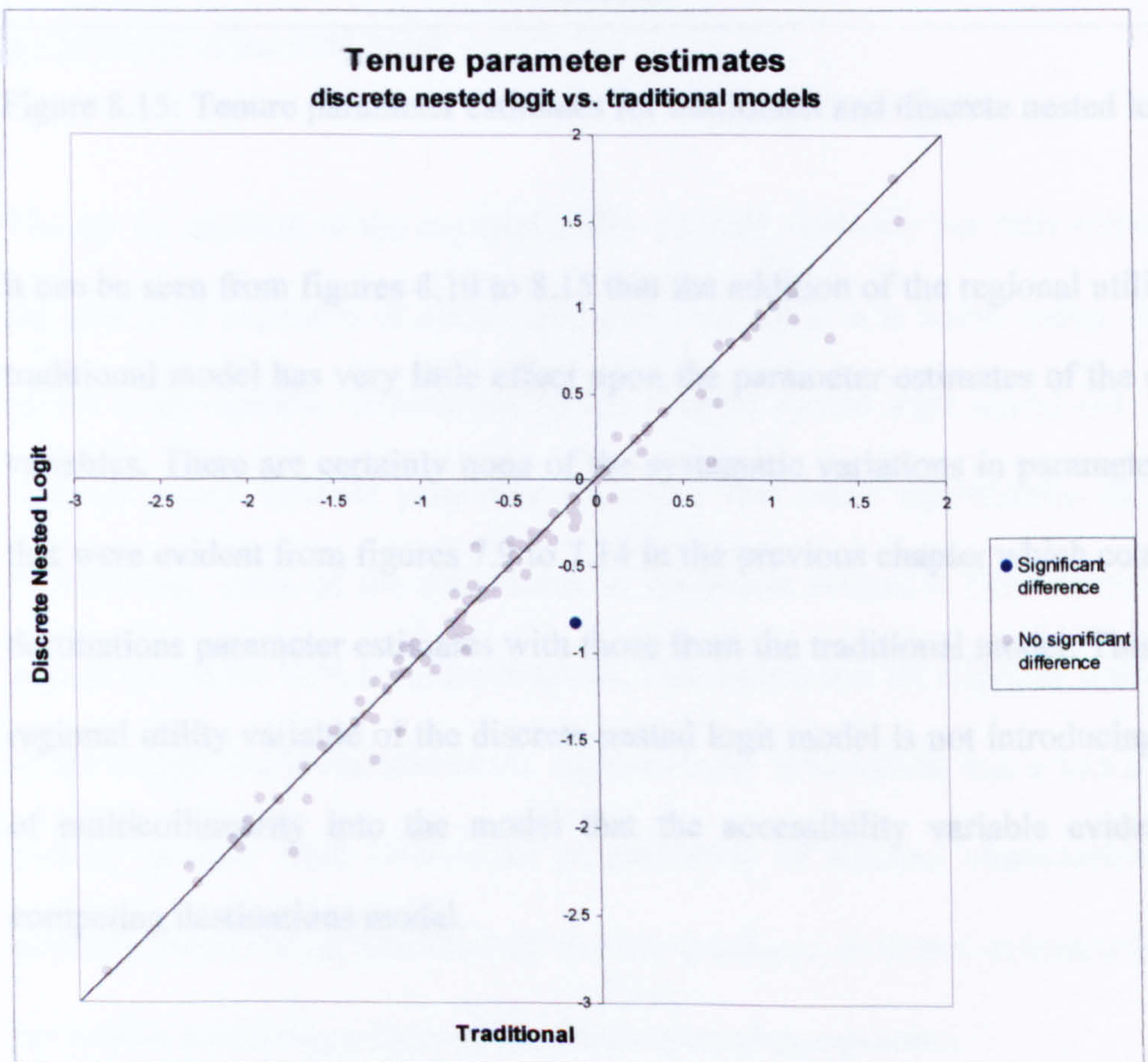


Figure 8.14: Tenure parameter estimates, traditional and discrete nested logit models.

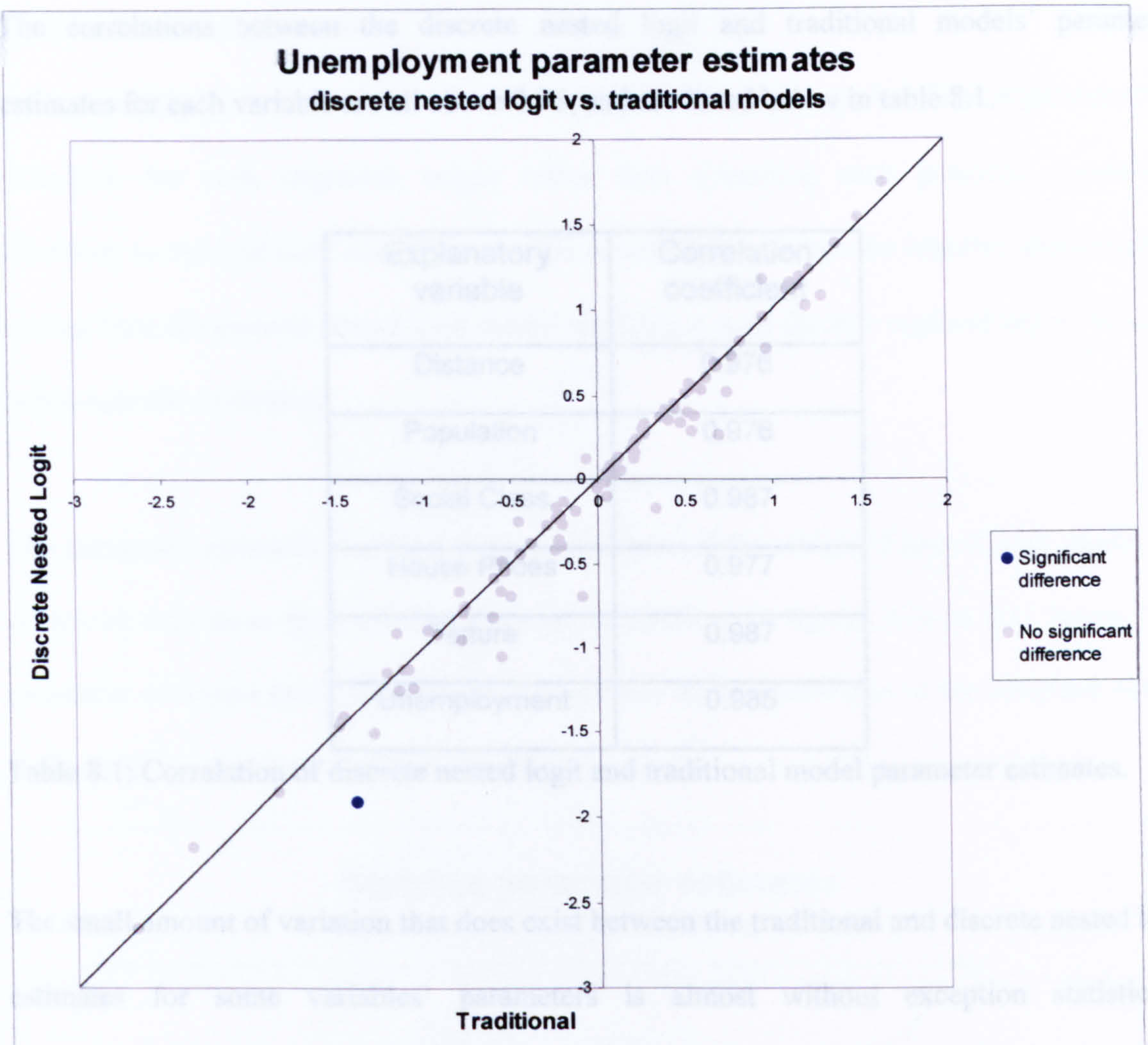


Figure 8.15: Tenure parameter estimates for traditional and discrete nested logit models.

It can be seen from figures 8.10 to 8.15 that the addition of the regional utility variable to the traditional model has very little effect upon the parameter estimates of the other explanatory variables. There are certainly none of the systematic variations in parameter estimate values that were evident from figures 7.9 to 7.14 in the previous chapter which compared competing destinations parameter estimates with those from the traditional model. This suggests that the regional utility variable of the discrete nested logit model is not introducing the same degree of multicollinearity into the model that the accessibility variable evidently did into the competing destinations model.

The correlations between the discrete nested logit and traditional models' parameter estimates for each variable are all above 0.96, and are listed below in table 8.1.

Explanatory variable	Correlation coefficient
Distance	0.976
Population	0.976
Social Class	0.987
House Prices	0.977
Tenure	0.987
Unemployment	0.985

Table 8.1: Correlation of discrete nested logit and traditional model parameter estimates.

The small amount of variation that does exist between the traditional and discrete nested logit estimates for some variables' parameters is almost without exception statistically insignificant at the 95% level.

Though the addition of the regional utility variable evidently has little systematic effect upon the parameter estimates of most explanatory variables, it is worth noting that more than half of the origin-specific calibrations of the discrete nested logit model (54 of 100) produced regional utility variable parameter estimates that were significantly different from zero. Furthermore, many of the differences in parameter estimates plotted above are statistically significant (at the 95% confidence level). This means that for migrants leaving more than half of the origins under consideration, regional-level information was a factor in their decision-making process. Such widespread consideration of regional characteristics as well as the generally improved goodness-of-fit that this produces, is further evidence that many migrants are indeed employing a hierarchical destination choice process.

Weighted nested logit model

As mentioned above, the weighted nested logit model employs a probabilistic regionalization definition for each migration origin rather than allocating each potential destination discretely to one and only one regional. This approach overcomes the intuitive shortcomings arising from the discrete nested logit model requiring a rigid discrete regionalization for each origin-specific calibration.

The parameter estimates resulting from calibrations of the weighted nested logit model are compared with those from traditional model calibrations in figures 8.16 to 8.21 below. The parameter estimates (and goodness-of-fit statistics) from calibrations of the weighted nested logit model are tabulated in full in appendix G.

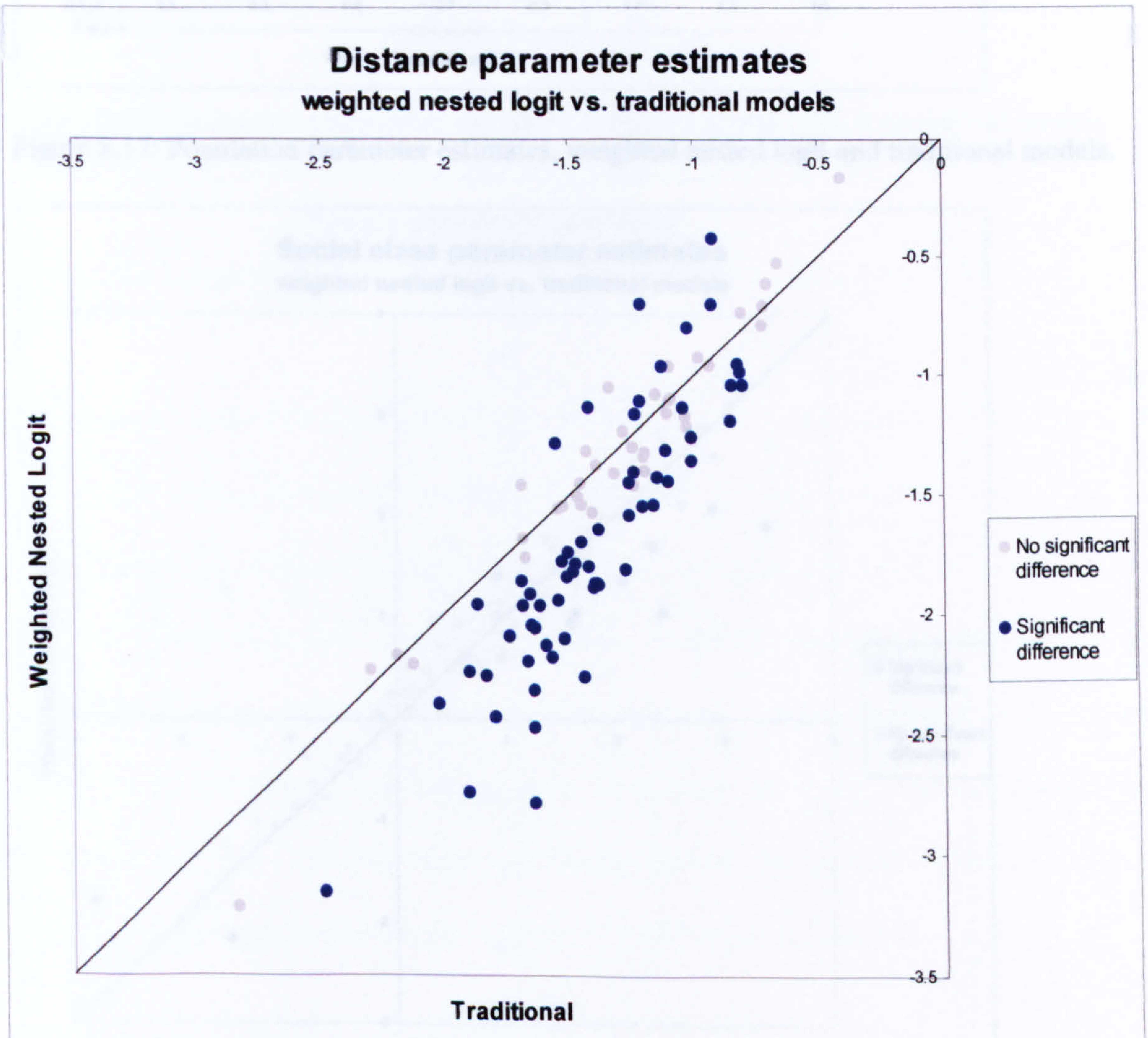


Figure 8.16: Distance parameter estimates, weighted nested logit and traditional models.

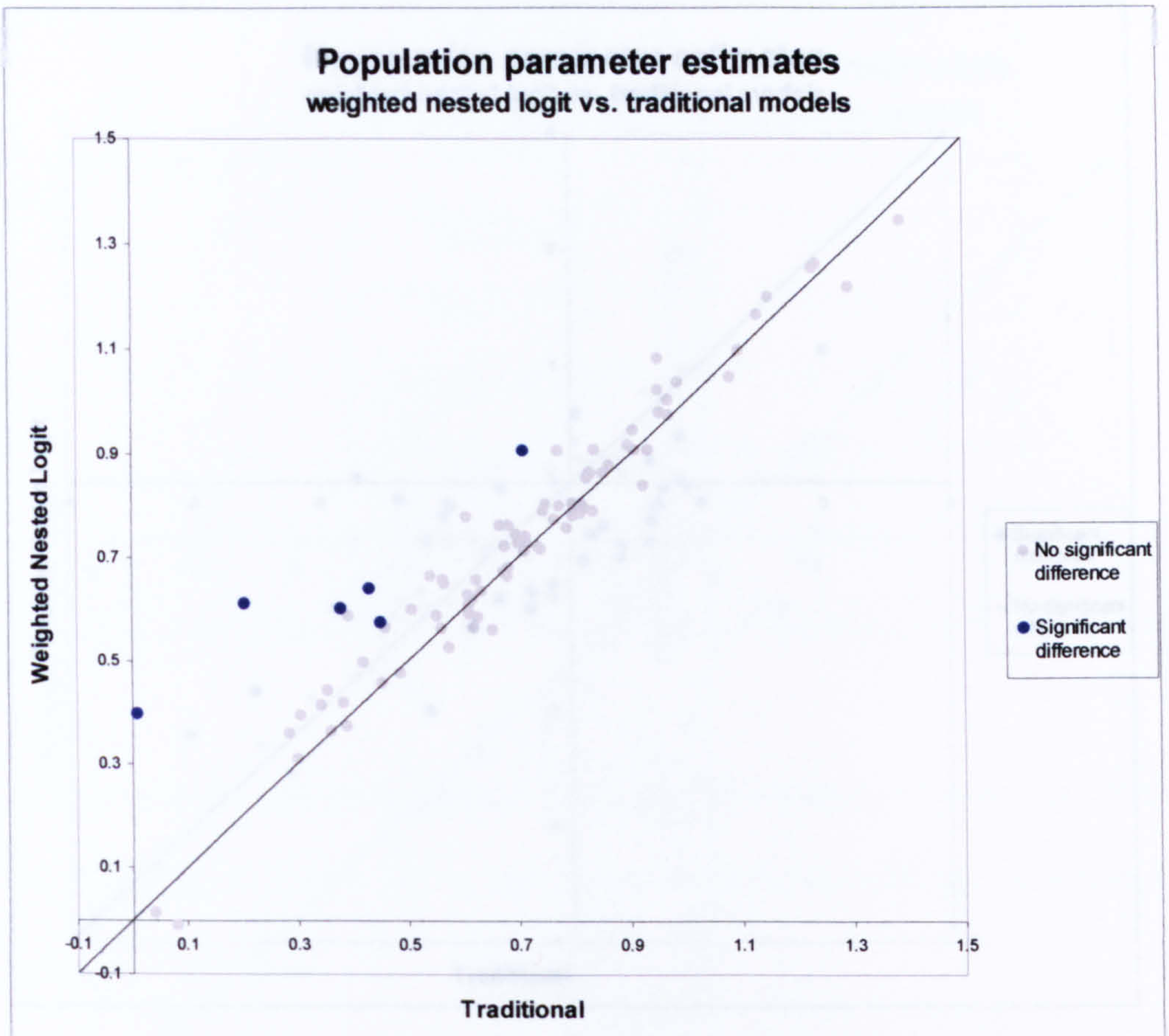


Figure 8.17: Population parameter estimates, weighted nested logit and traditional models.

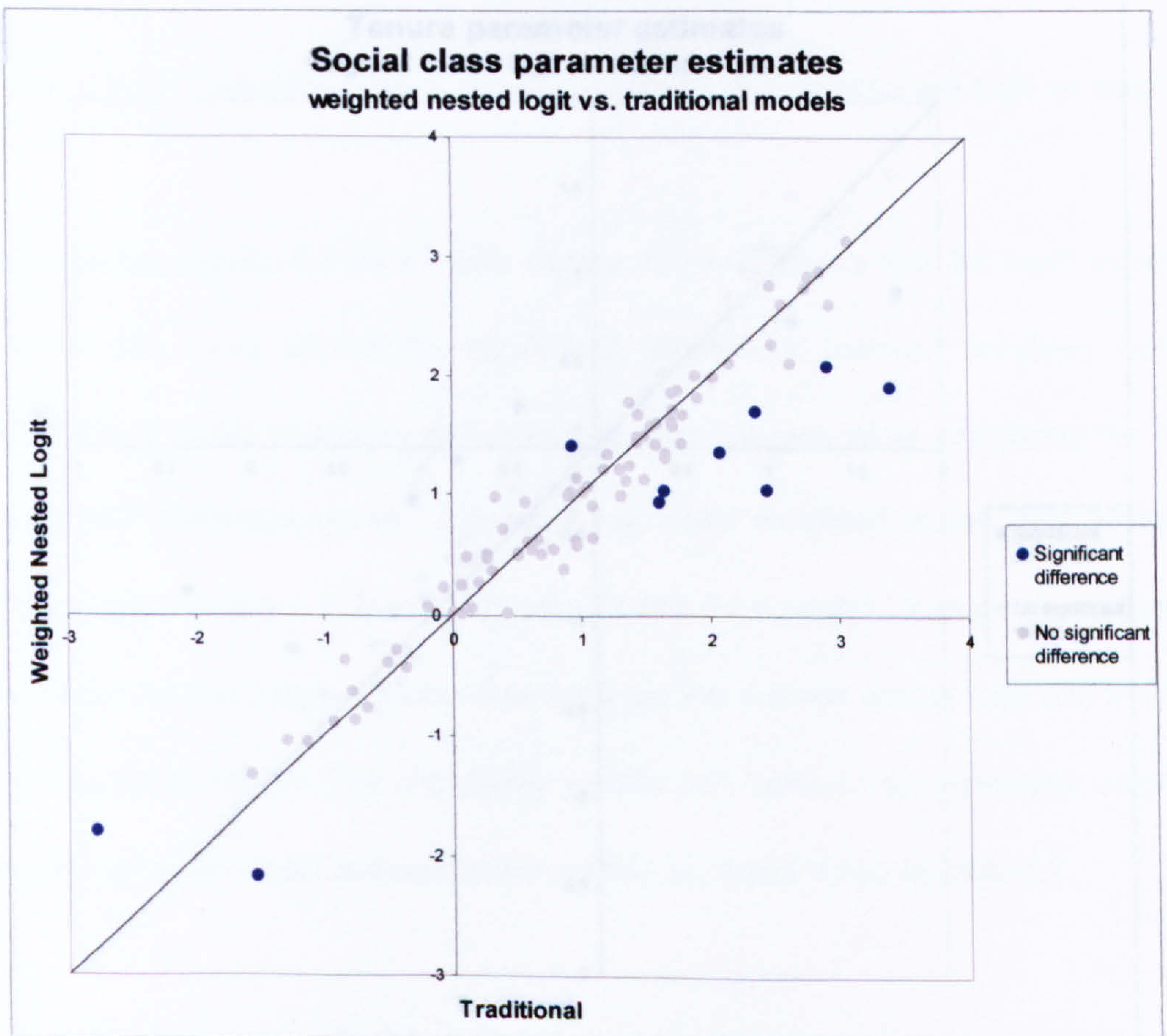


Figure 8.18: Social class parameter estimates, weighted nested logit & traditional models.

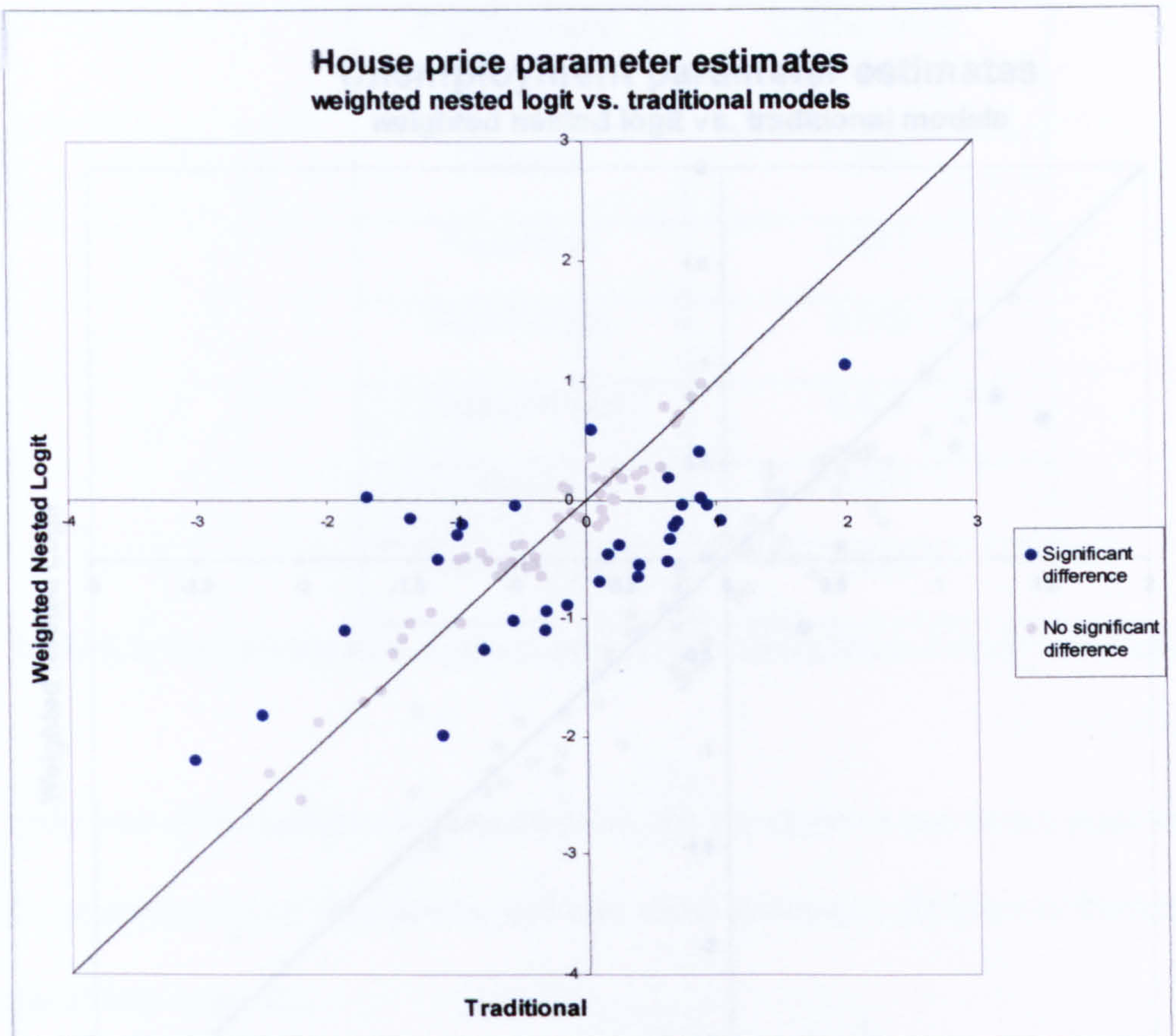


Figure 8.19: House price parameter estimates, weighted nested logit & traditional models.

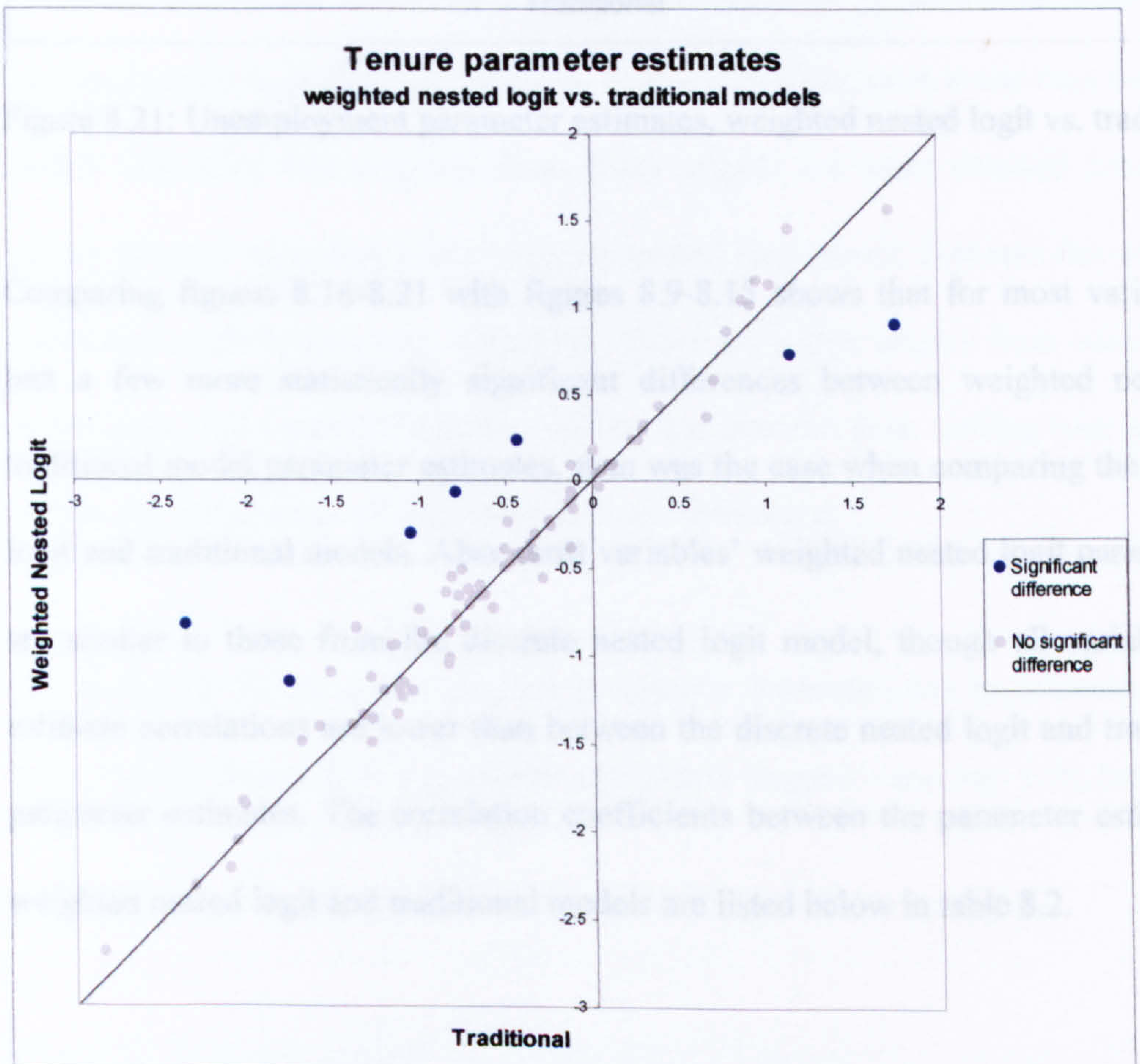


Figure 8.20: Tenure parameter estimates, weighted nested logit and traditional models.

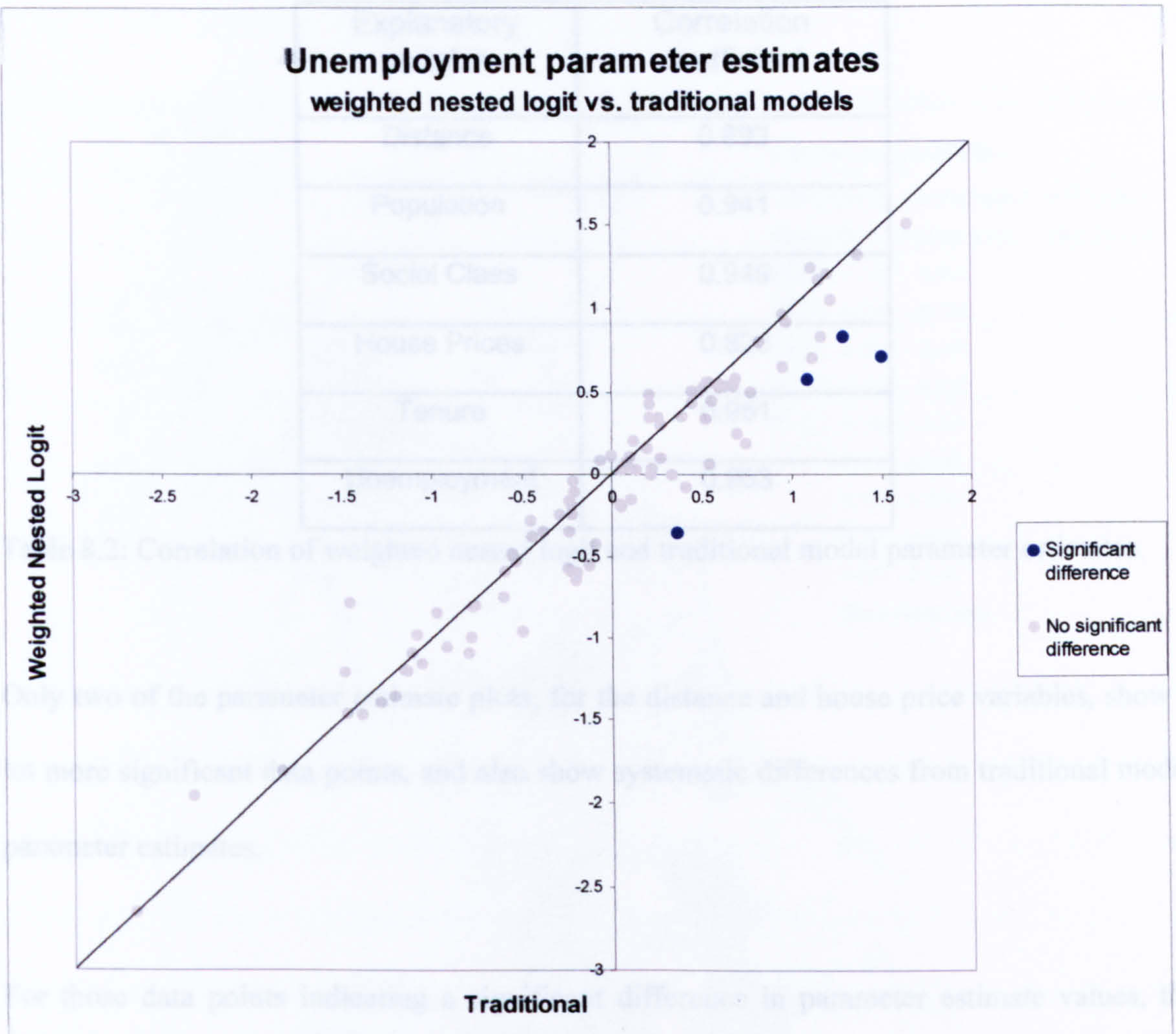


Figure 8.21: Unemployment parameter estimates, weighted nested logit vs. traditional.

Comparing figures 8.16-8.21 with figures 8.9-8.15 shows that for most variables there are just a few more statistically significant differences between weighted nested logit and traditional model parameter estimates, than was the case when comparing the discrete nested logit and traditional models. Also, most variables' weighted nested logit parameter estimates are similar to those from the discrete nested logit model, though all variables' parameter estimate correlations are lower than between the discrete nested logit and traditional models parameter estimates. The correlation coefficients between the parameter estimates from the weighted nested logit and traditional models are listed below in table 8.2.

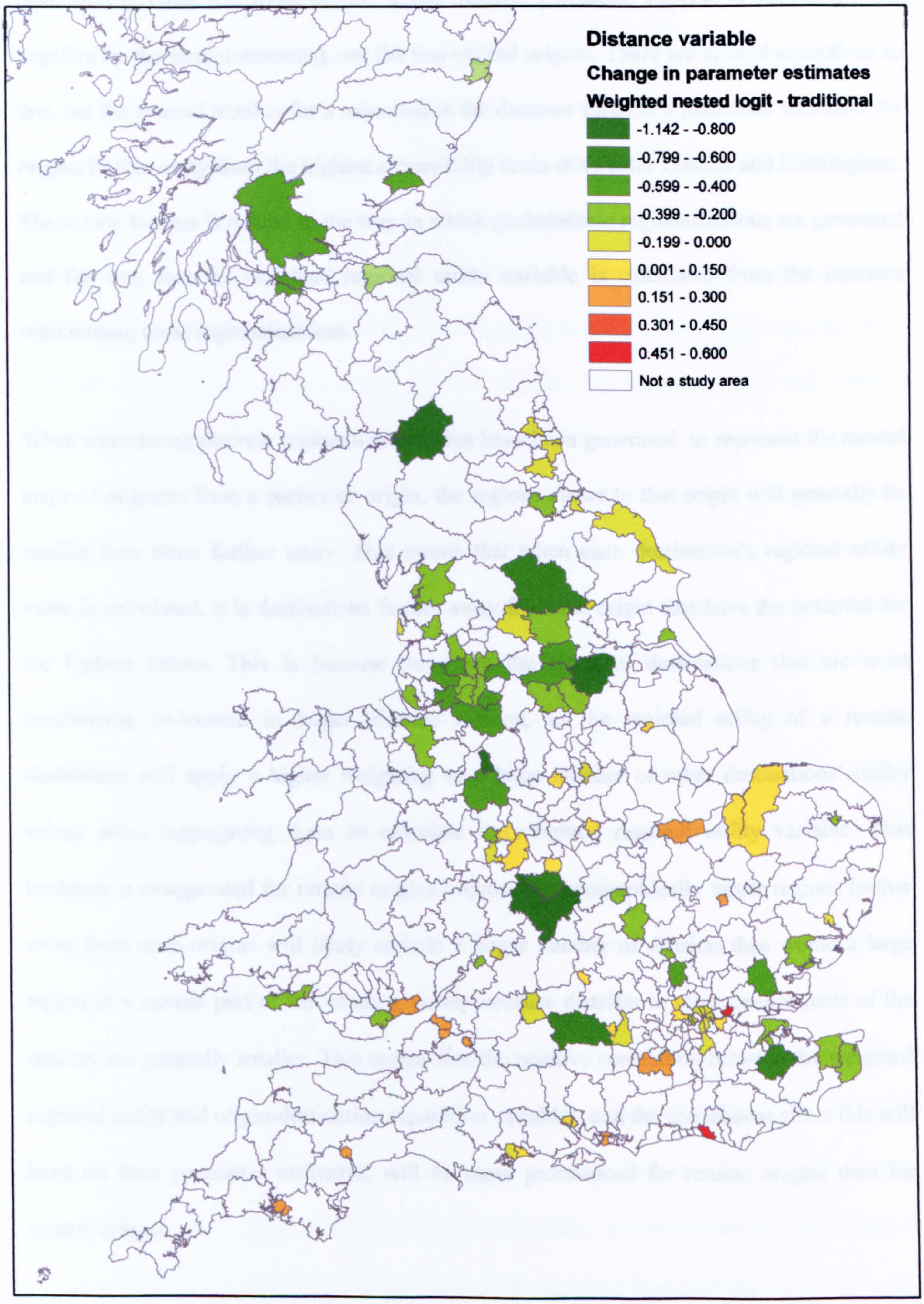
Explanatory variable	Correlation coefficient
Distance	0.893
Population	0.941
Social Class	0.949
House Prices	0.826
Tenure	0.951
Unemployment	0.963

Table 8.2: Correlation of weighted nested logit and traditional model parameter estimates.

Only two of the parameter estimate plots, for the distance and house price variables, show a lot more significant data points, and also show systematic differences from traditional model parameter estimates.

For those data points indicating a significant difference in parameter estimate values, the weighted nested logit distance parameter estimates are generally lower than for the traditional model, indicating that migrants from many origins are more deterred from moving over greater distance according to the weighted nested logit model, than was the case based on the traditional model's parameter estimates. There are a few origins from which the weighted nested logit model predicts migrants will be less deterred from moving over longer distances, but the key trend is one of increasing distance deterrence.

The reason for this selective increase in distance deterrence can be understood if the spatial pattern of the change in this parameter estimate is mapped – see map 8.10, below.



Map 8.10: Distance parameter estimates, weighted nested logit – traditional model.

It can be seen from map 8.10 that those origins that are seeing the largest increase in their distance deterrence (i.e. those origins whose distance parameter estimate is becoming more negative by the largest amounts), are the less central origins. There are several exceptions to this, but the general trend is for a reduction in the distance variable's parameter estimates for origins further away from the highest accessibility areas of Greater London and Birmingham. The reason for this is related to the way in which probabilistic regionalizations are generated and the way that the weighted regional utility variable is calculated from the matrices representing those regionalizations.

When considering discrete regionalizations that have been generated to represent the mental maps of migrants from a particular origin, the regions closer to that origin will generally be smaller than those further away. This means that when each destination's regional utility value is calculated, it is destinations further away from the origin that have the potential for the highest values. This is because those are the types of destinations that are most consistently co-located in larger discrete regions, so the regional utility of a remote destination will apply a higher weighting to a large number of other destinations' utility values when aggregating them to calculate its weighted regional utility variable. This tendency is exaggerated for remote origins because the geographically larger regions further away from such origins will likely contain a larger number of districts than would a large region in a remote part of the country, purely because districts in more central parts of the country are generally smaller. This means that the positive correlation between the weighted regional utility and origin-destination separation variables, and the consequent effect this will have on their parameter estimates, will be more pronounced for remote origins than for central origins.

A marked downward shift in house price parameter estimates is also evident from figure 8.19, for a subset of migrant origins. This could be a reflection of the spatial scale at which house prices affect migration destination choice behaviour.

There will inevitably be some interplay between the destination specific house price variable and the regional utility variable, because surrounding areas' house prices are one of the variables contributing to the value of the regional utility variable. Also, intuitively, house prices are one of the key factors that migrants are likely to consider at a wider regional level - migrants from origins in the north of the country, for instance, may well make a generalization that they cannot afford to move to Greater London or to the South East in general, because house prices are so high down there. For migrants from more expensive and affluent areas, house price may be less important at the regional level, because it does not act to remove whole regions from their choice set (unless of course low house prices deter such migrants), but it may continue to act as a significant variable when distinguishing between neighbouring destinations at a smaller spatial scale within a selected region. For such migrants, any multicollinearity introduced due to correlation between destinations regional utility variable and its house price could cause the observed offset in the house price parameter estimates from a subset of migration origins.

Hybrid Nested Logit Model

The parameter estimates predicted by the hybrid weighted nested logit model are compared with those predicted by the traditional model in figures 8.22 through 8.27, below. Once again, a distinction is made between those pairs of parameter estimates that are significantly different from each other (at 95% level), plotted in blue, and those that are not, plotted in grey. The parameter estimates (and goodness-of-fit statistics) from calibrations of the hybrid weighted nested logit model are tabulated in full in appendix H.

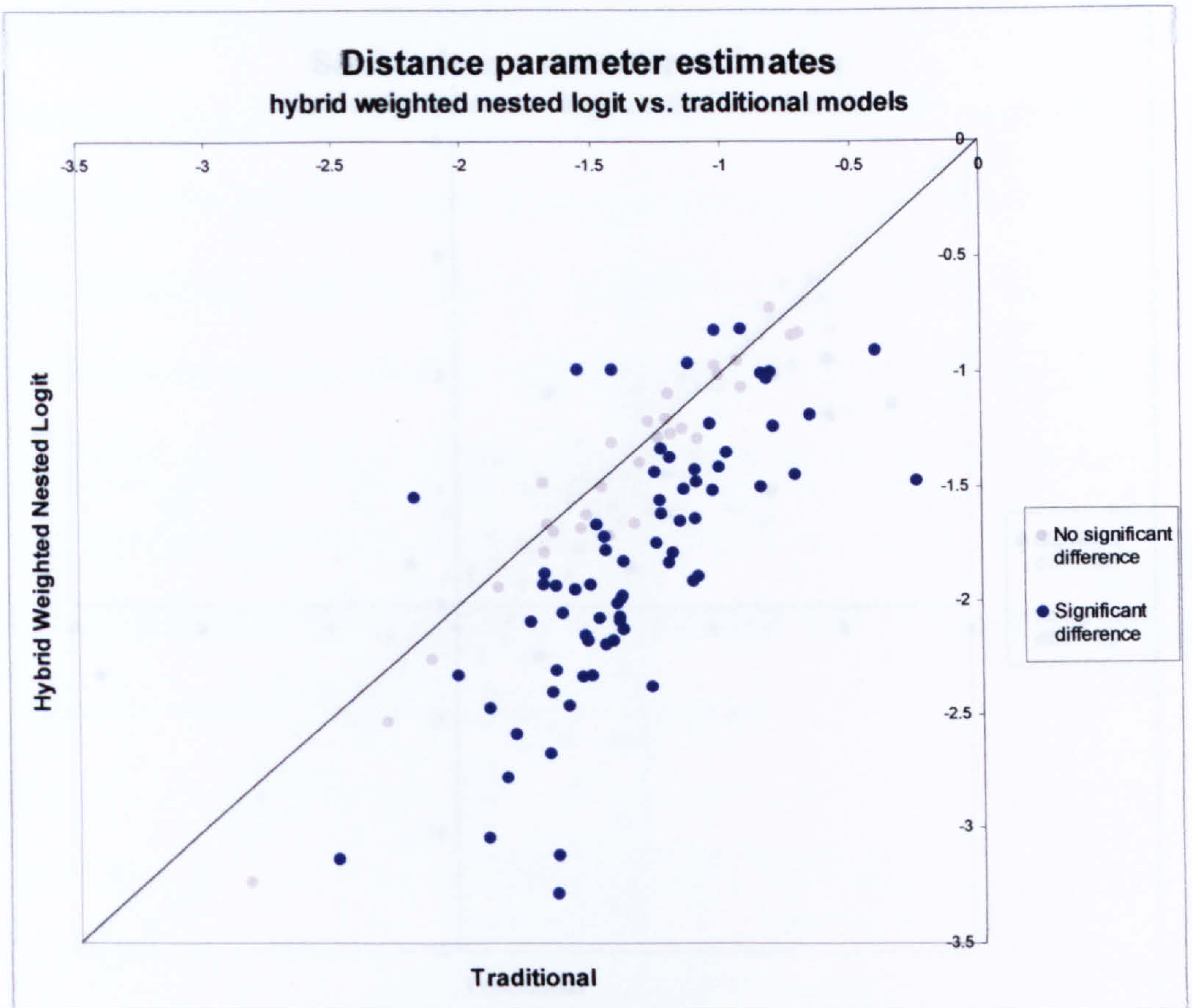


Figure 8.22: Distance parameter estimates, hybrid weighted nested logit vs. traditional.

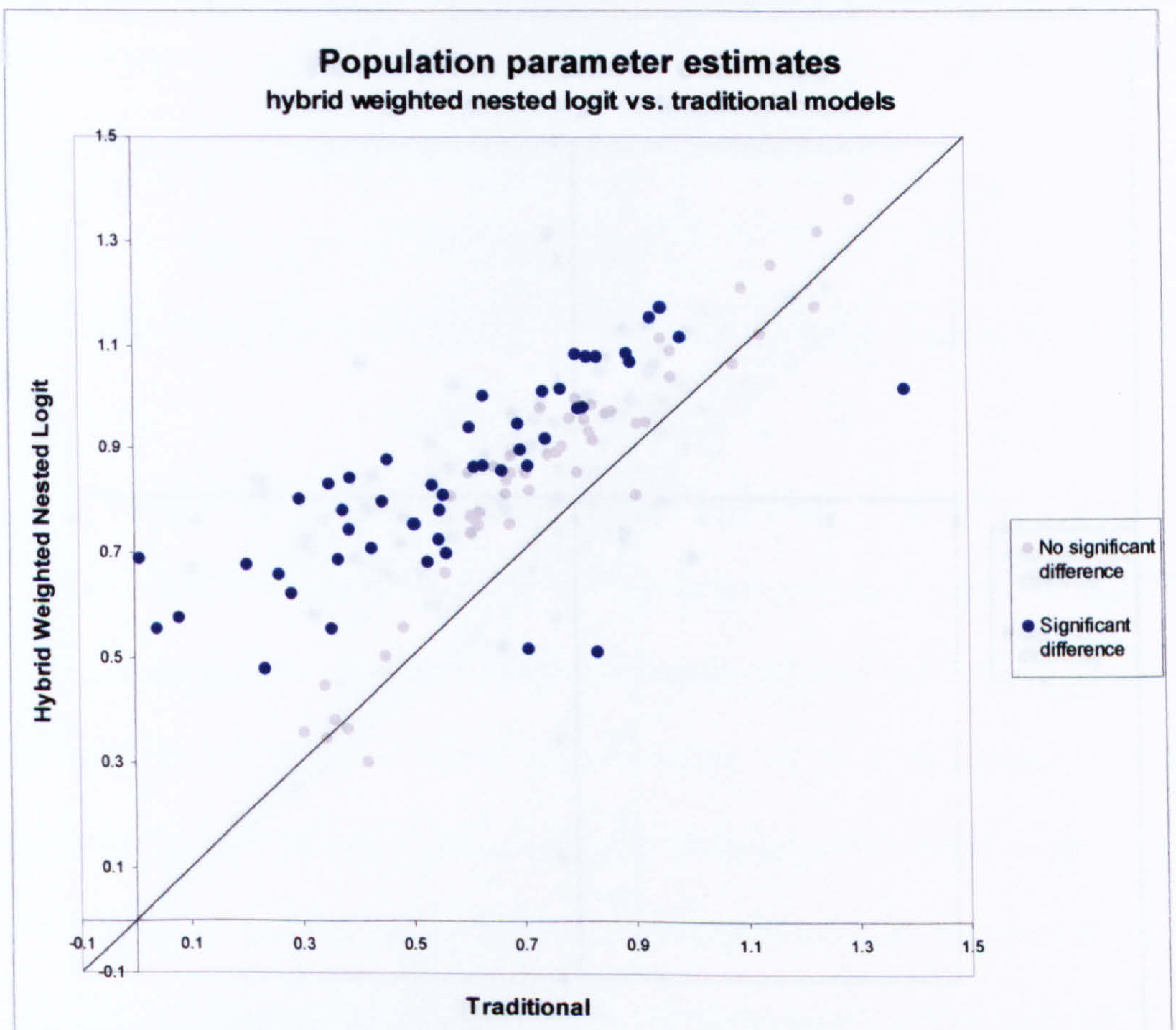


Figure 8.23: Population parameter estimates, hybrid weighted nested logit vs. traditional.

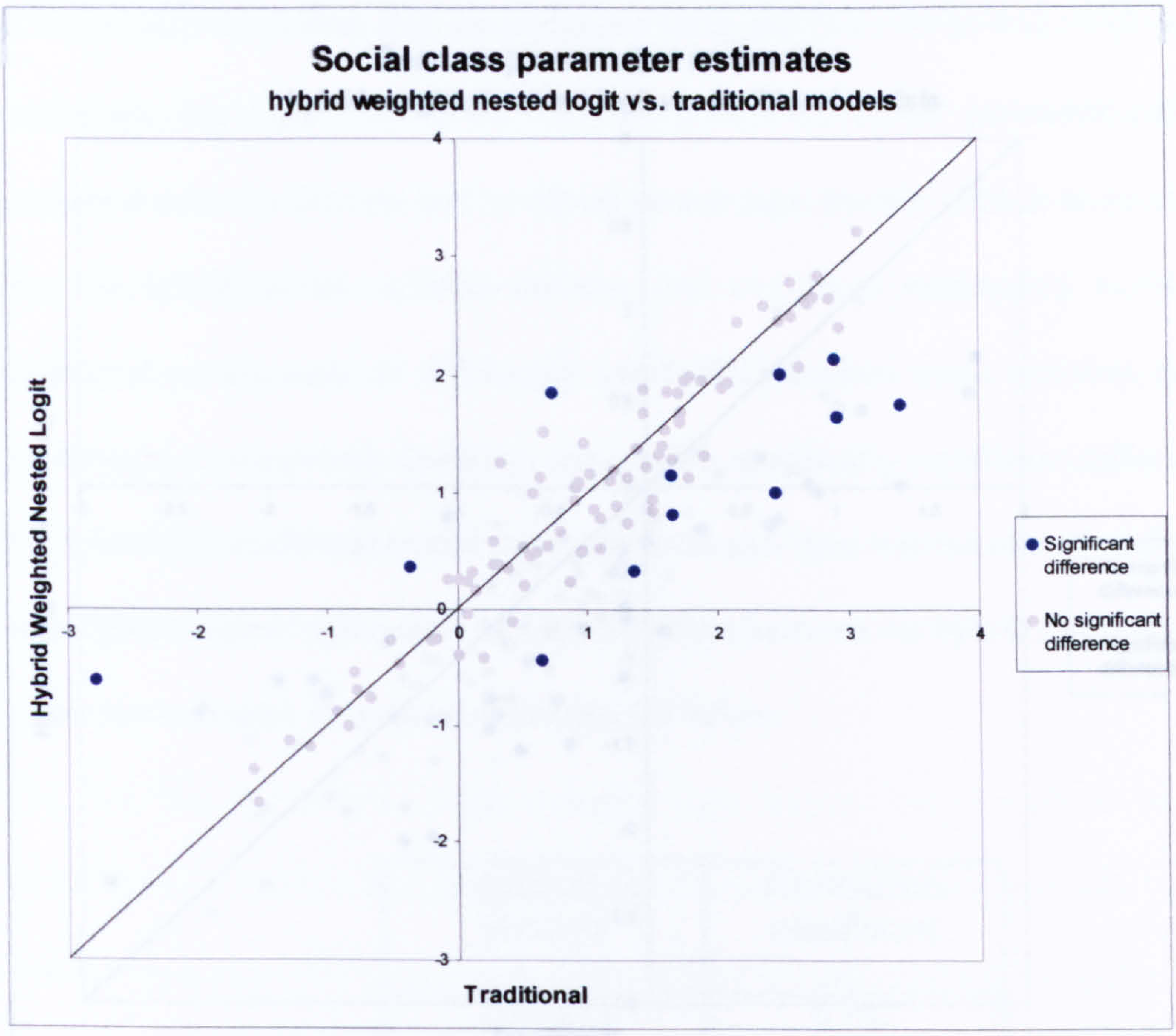


Figure 8.24: Social class parameter estimates, hybrid weighted nested logit vs. traditional.

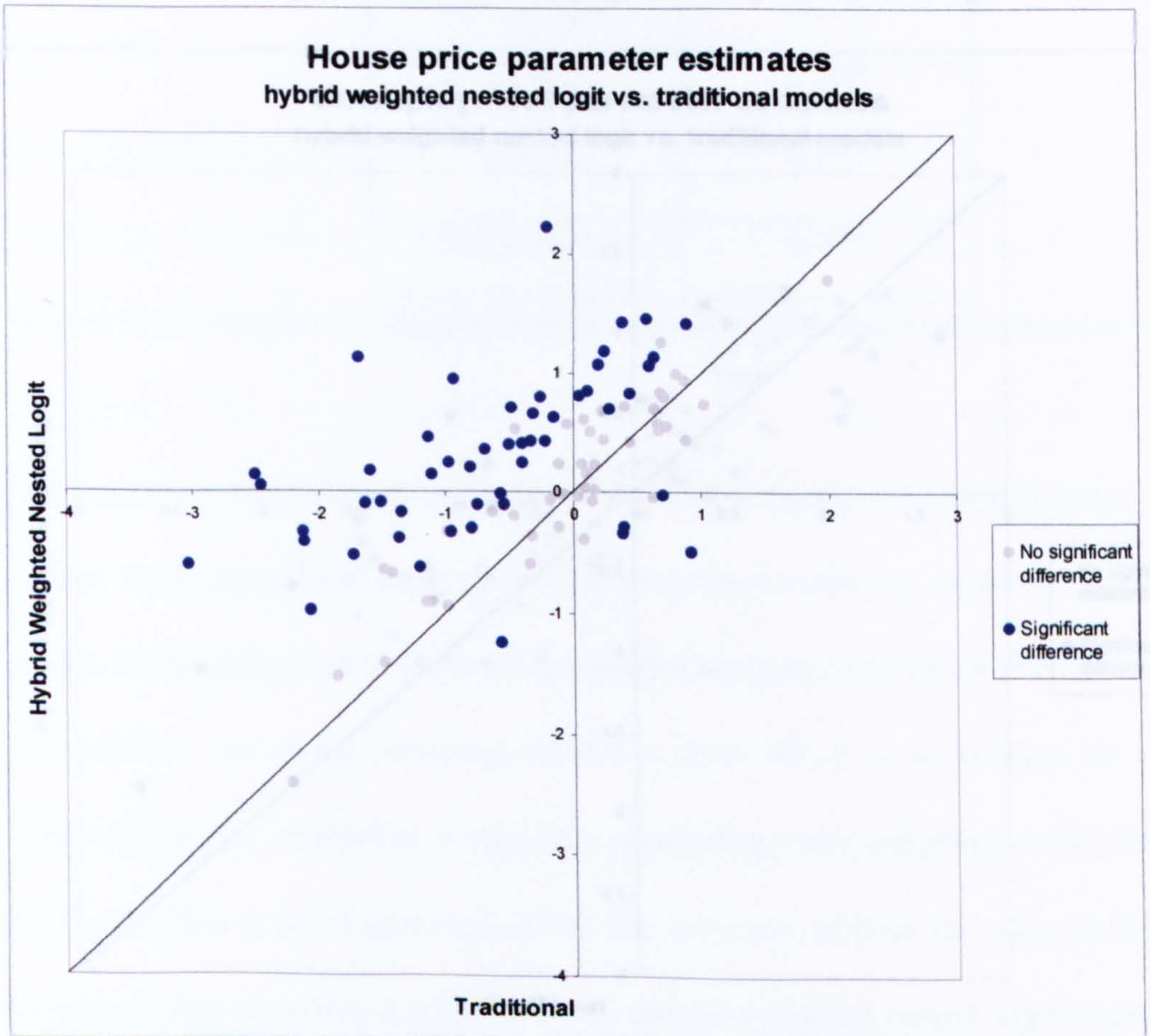


Figure 8.25: House price parameter estimates, hybrid weighted nested logit vs. traditional.

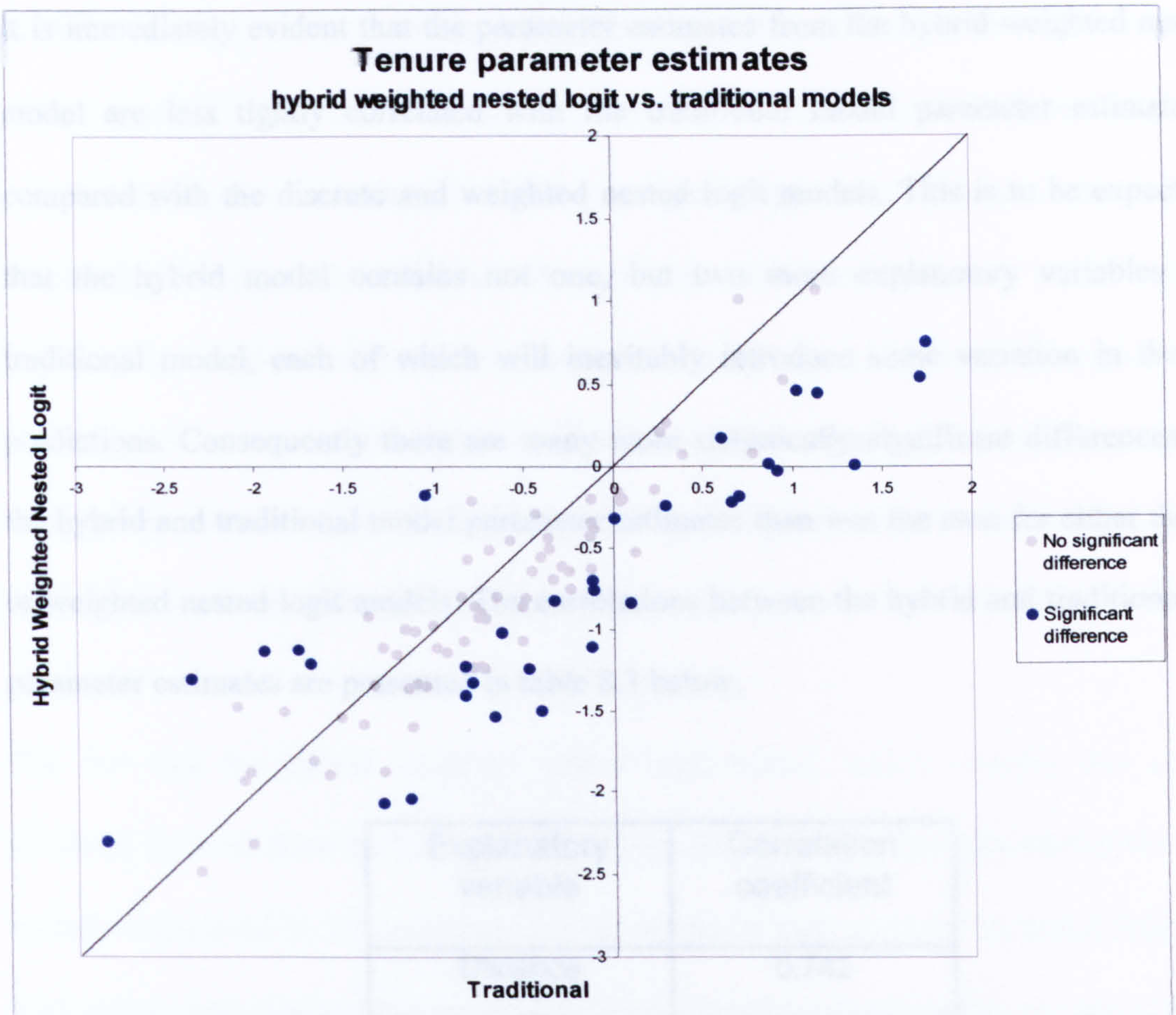


Figure 8.26: Tenure parameter estimates, hybrid weighted nested logit vs. traditional.

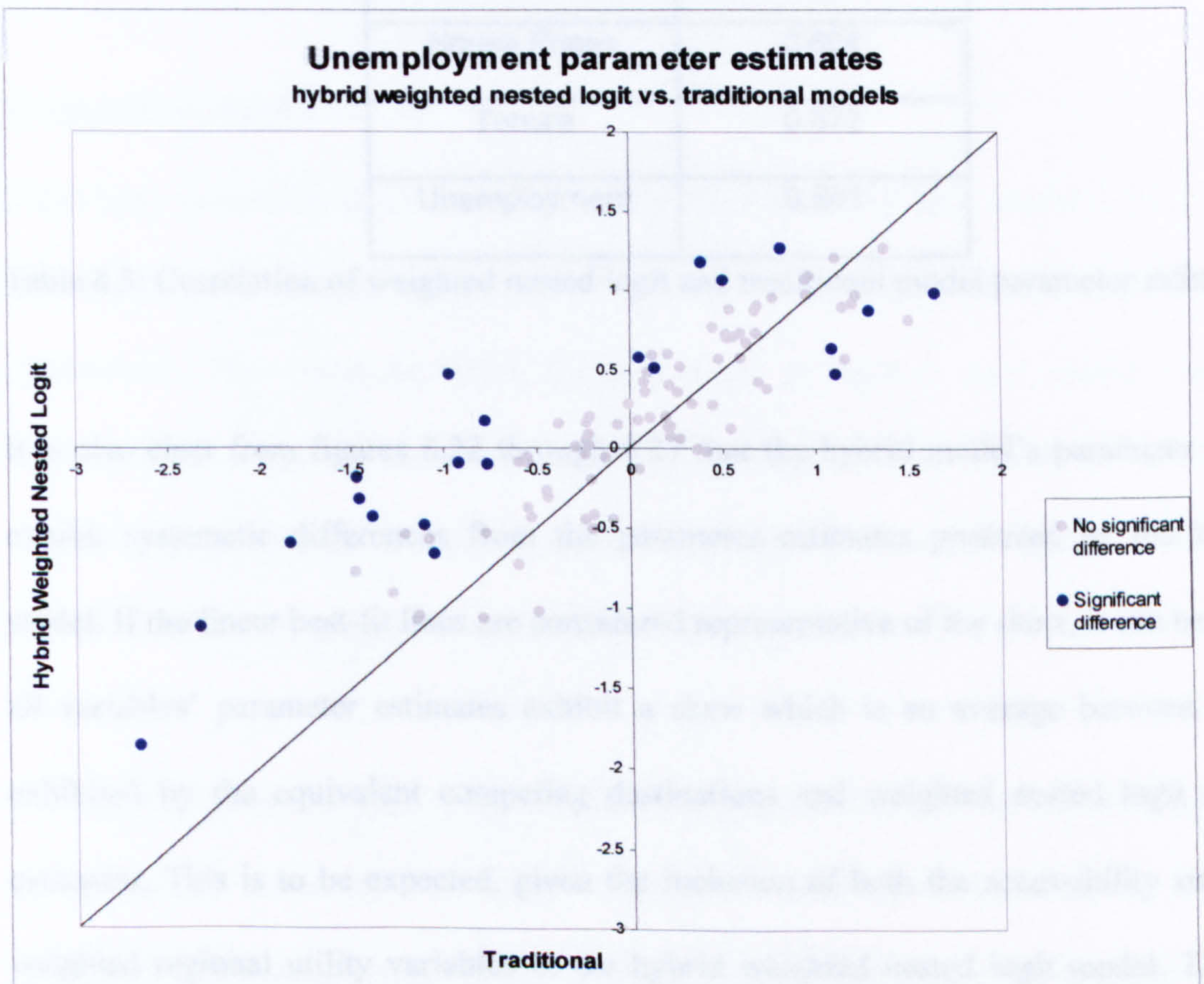


Figure 8.27: Unemployment parameter estimates, hybrid weighted nested logit vs. traditional.

It is immediately evident that the parameter estimates from the hybrid weighted nested logit model are less tightly correlated with the traditional model parameter estimates, when compared with the discrete and weighted nested logit models. This is to be expected given that the hybrid model contains not one, but two more explanatory variables than the traditional model, each of which will inevitably introduce some variation in the model's predictions. Consequently there are many more statistically significant differences between the hybrid and traditional model parameter estimates than was the case for either the discrete or weighted nested logit models. The correlations between the hybrid and traditional models' parameter estimates are presented in table 8.3 below.

Explanatory variable	Correlation coefficient
Distance	0.742
Population	0.803
Social Class	0.904
House Prices	0.604
Tenure	0.872
Unemployment	0.861

Table 8.3: Correlation of weighted nested logit and traditional model parameter estimates.

It is also clear from figures 8.22 through 8.27 that the hybrid model's parameter estimates exhibit systematic differences from the parameter estimates produced by the traditional model. If the linear best-fit lines are considered representative of the skew, it can be seen that all variables' parameter estimates exhibit a skew which is an average between the skew exhibited by the equivalent competing destinations and weighted nested logit parameter estimates. This is to be expected, given the inclusion of both the accessibility statistic and weighted regional utility variables in the hybrid weighted nested logit model. Thus, these systematic differences between the parameter estimates from the hybrid and traditional models very likely arises from the multicollinearity effects that have been discussed at length above, some of which operate generally, affecting all origins, and other in a more specific way because of a correlation between origin-destination separation and one of the hierarchical explanatory variables. However, as with the competing destinations model discussed in chapter 7, the author contends that the marked improvement in goodness-of-fit of the hybrid model over the traditional and weighted nested models demonstrates that the explanatory power of the accessibility variable outweighs the limited multicollinearity that its inclusion introduces to the hybrid nested logit model.

The fact that the hybrid weighted nested logit model, which contains two explanatory variables derived from hierarchical principles, provides the best goodness-of-fit of all the models considered in this research, further suggests that migrants do group their potential destinations into clusters or regions and consider some attributes at this aggregate level when deciding on their migration destinations.

Regionalization sensitivity

It has been mentioned that a weakness of the discrete nested logit model is the fact that it must be calibrated within the context of a single discrete regionalization of all potential destinations. The extent to which the imposition of such a rigid spatial hierarchy is detrimental to migration destination choice modelling can be shown by examining the sensitivity of the discrete nested logit model's goodness-of-fit and parameter estimates to the specific discrete regionalization against which it is calibrated.

Leeds was selected as the migrant origin for examination of the nested logit models' regionalization sensitivity because R^2 results show that the weighted nested logit model performs significantly better than the discrete nested logit model for migration from this origin (weighted nested logit model gives a 5% higher R^2). This improvement could result from the fact that the 'best' discrete regionalisation used to calibrate the discrete nested logit model for this origin was atypical. Notably, Leeds also has high out-migration which produces parameter estimates with lower standard error – indeed all parameter estimates for model calibrations were statistically significantly different from zero for migration from Leeds. Maps 8.3 and 8.4 below show the top two discrete regionalizations that were generated to represent the spatial structure cognized by migrants leaving Leeds. Note that whilst both of these regionalizations have similarly low variances in regional information (as defined and described in chapter 6) they are nonetheless noticeably different in spatial structure.

Table 8.4, below, presents the parameter estimates and goodness-of-fit statistics from calibrations of the discrete nested logit model using the two 'best' discrete regionalizations for migrants from Leeds (shown previously in maps 6.3 and 6.4).

	Regionalization 1		Regionalization 2	
Parameter name	Parameter estimate	Standard error	Parameter estimate	Standard error
Distance	-1.49	0.018	-1.47	0.018
Population	1.08	0.035	1.05	0.036
Social class	*0.08*	*0.123*	*-0.02*	*0.124*
House prices	0.43	0.092	0.63	0.088
Tenure	-1.84	0.093	-1.79	0.096
Unemployment	-1.25	0.082	-1.17	0.085
Regional utility	0.48	0.063	0.26	0.072
Adjusted R^2	0.928		0.937	

Table 8.4: Discrete nested logit results for Leeds using two different regionalizations.

It can be seen from table 8.4 that reasonably large differences exist between the results obtained from these two calibrations of the discrete nested logit model. Owing to the semi-random manner in which discrete regionalizations have been generated for the purposes of this research, neither of these calibrations can be said to be more 'correct' that the other. Furthermore, even if a deterministic regionalization algorithm were used to construct the regionalization, in real terms any 'mathematically optimal' regionalization produced by such an algorithm would be likely to be no more accurate a representation of any particular migrant's mental maps than either of the discrete regionalizations shown above. The problem lies not in the precise means of regionalization generation, but in the fact that only one regionalization is used to approximate the mental maps of a large group of heterogeneous migrants. This is an inherently flawed and over-simplistic approach.

As discussed above, the discrete nested logit model was modified to produce the weighted nested logit model which overcomes the over-simplification of using a single discrete regionalization by employing a regional utility variable that is based on a probabilistic rather than a discrete definition of regions. Such a probabilistic regionalization is represented as a matrix, any specific element of which represents the likelihood of two specific destinations being perceived as being in the same region by a migrant from a particular origin. This is an intuitively more acceptable model, and due to the manner in which probabilistic regionalizations are generated (described in chapter 6), it also exhibits much lower sensitivity to calibration against different regionalizations.

In order to assess the sensitivity of the weighted nested logit model to the specific probabilistic regionalization that is used in its calibration, two probabilistic regionalizations were independently generated for the same migration origin, Leeds, from two separate runs of the regionalization algorithm. Each of these runs generated 10,000 valid discrete regionalizations for migrants leaving Leeds and then aggregated a probabilistic regionalization matrix from the 'best' 1000 of these regionalizations (as ranked by their regional information variance – lower being better).

It is very hard to effectively visualize such a large and interconnected dataset as a probabilistic regionalization matrix in a meaningful way. For instance, a connectivity map with 459x459 connections would be virtually impossible to interpret. It is easier to map the likely regional coexistence of a specific destination, but that approach requires 459 maps in order to represent the information in a single origin-specific probabilistic regionalisation matrix. However, the similarity between the two regionalizations is evident from the extremely high correlation coefficient between the two matrices: 0.997. Also, when the weighted nested logit model is calibrated using the two different regionalizations, the resulting R^2_{adj} values are identical, as one might expect.

The parameter estimates resulting from the two calibrations of the weighted nested logit model for migration from Leeds are shown in table 8.5 below.

	Regionalization 1		Regionalization 2	
Parameter name	Parameter estimate	Standard error	Parameter estimate	Standard error
Distance	-1.70	0.026	-1.69	0.025
Population	1.05	0.036	1.03	0.036
Social Class	0.29	0.125	0.33	0.126
House Prices	-0.04	0.101	-0.05	0.099
Tenure	-1.43	0.091	-1.39	0.093
Unemployment	-1.16	0.079	-1.14	0.079
Weighted regional utility	-1.77	0.135	-1.80	0.127
R^2_{adj}	0.940		0.941	

Table 8.5: Weighted nested logit results for Leeds using two different regionalizations.

The sensitivity of the weighted nested logit parameter estimates (see table 8.5) can be seen to be lower than was the case for the discrete weighted logit model, see table 8.4. runs calibrated against different discrete regionalizations.

Another way in which the regionalization sensitivity of the discrete nested logit model can be demonstrated is to calibrate a set of 100 origin-specific model calibrations using the same discrete regionalization. Recall that the regionalization is generated in an origin-specific manner, such that a different discrete regionalization is normally used when calibrating the model for each different migration origin. Given the sensitivity demonstrated above of the discrete nested logit model's to the specific regionalization against which the model is calibrated, one would expect that when all origins' migration is predicted with respect to the same discrete regionalization, the performance of the model for the majority of origins would be lower than when each origin's migration is predicted with respect to a discrete regionalization that was created specifically for that origin.

Figures 8.28 and 8.29, below, compare the R^2_{adj} and AIC statistics resulting from two such sets of 100 origin-specific discrete nested logit model calibrations. One set of calibrations uses origin-specific discrete regionalizations and the other set imposes the same single discrete regionalization (originally created for the origin Aberdeen) on all 100 origins.

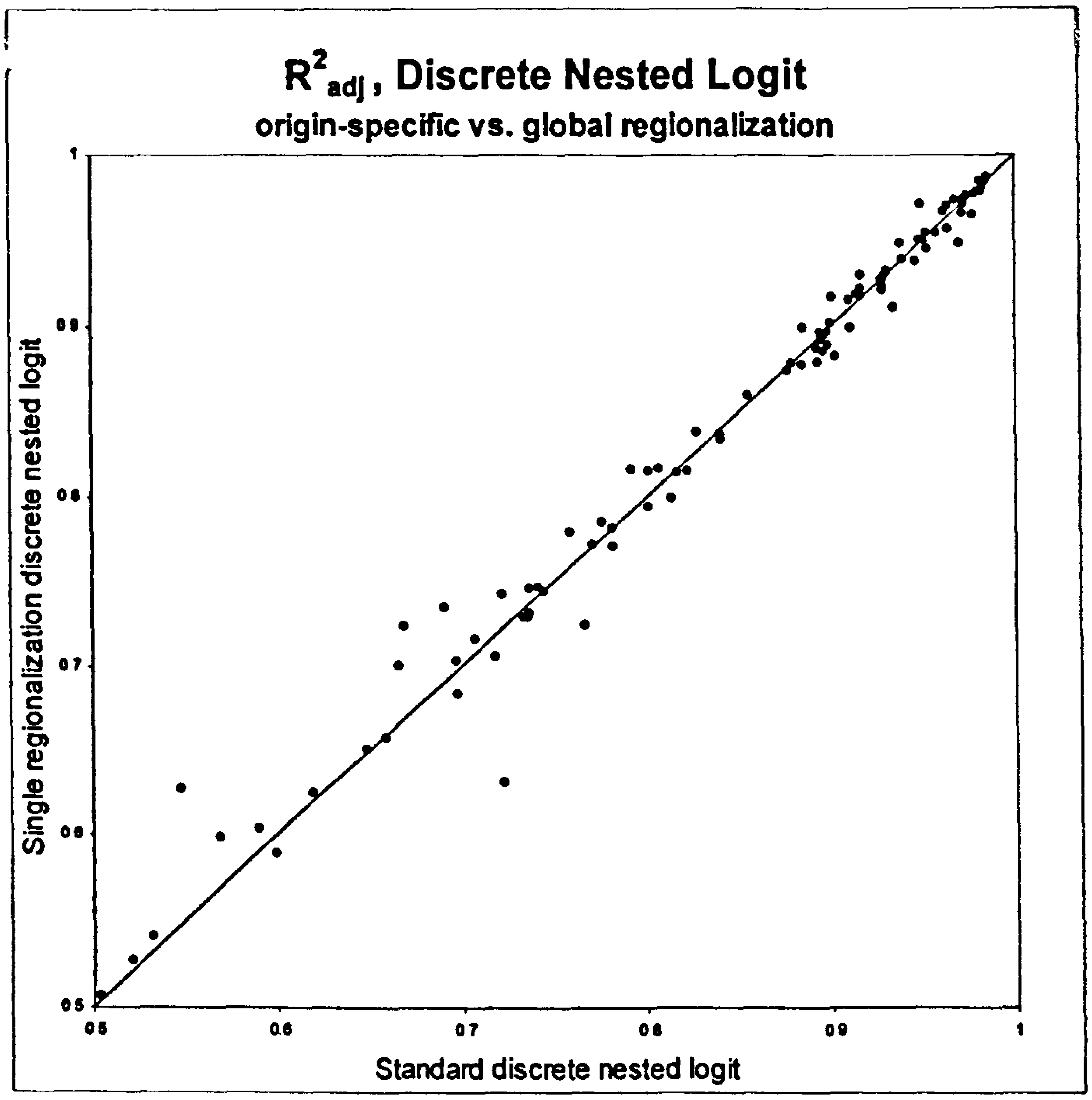


Figure 8.28: Discrete nested logit R^2_{adj} comparing origin-specific & global regionalizations.

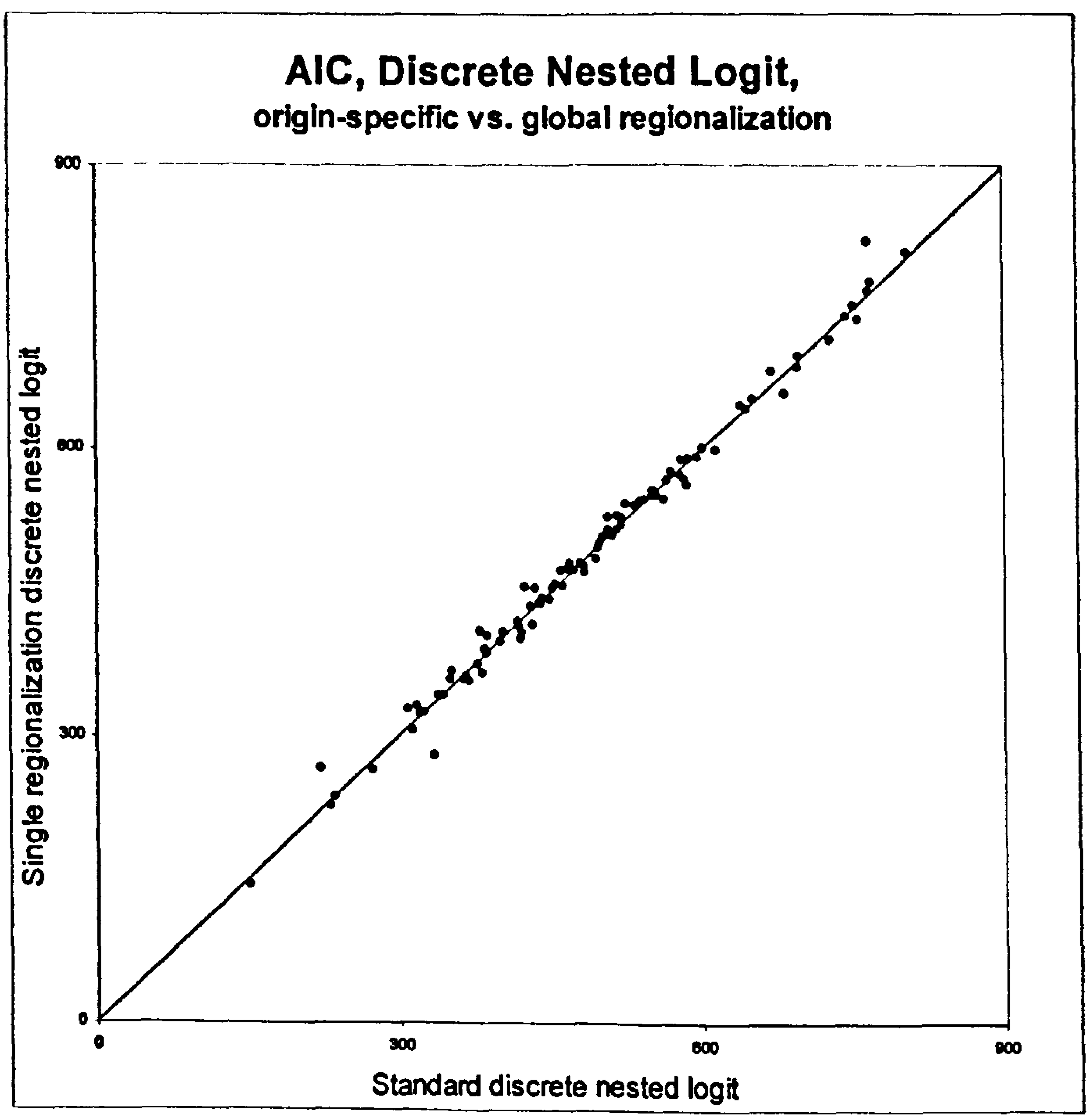


Figure 8.29: Discrete nested logit AIC comparing origin-specific & global regionalizations.

Figures 8.28 and 8.29 suggest that there is little or no benefit in generating origin-specific discrete regionalizations for use in calibrating discrete nested logit models. There are roughly as many origins for which perform is improved as reduced through the use of origin-specific regionalizations. This is not an intuitive finding, but recall that this is an adjusted R^2 statistic, so in those instances where the additional regional utility variable is of no value at all, the R^2_{adj} will likely reduce slightly because the reduction in the R^2_{adj} statistic due to the additional model complexity will not be offset by any improvement in its predictive ability. This accounts for those origins for which us of origin-specific regionalizations appear to have a detrimental effect. On balance one would expect that those origins which do appear to see a benefit from using origin-specific regionalizations will be those that exhibit positive regional utility parameter estimates. Indeed, there is a slight negative correlation (of -0.40) between the difference in AIC for origin-specific calibrations made using origin-specific regionalization minus those made with the same global regionalization, and the Z-score indicating the significance of the discrete nested logit model's regional utility variable. However, generally speaking, the results in figures 8.28 and 8.29 do not provide great confidence that origin-specific discrete regionalizations are providing worthwhile benefit compared to arbitrary use of any discrete regionalization.

The equivalent comparisons between use of origin-specific and global probabilistic regionalizations is presented in figures 8.30 and 8.31 below.

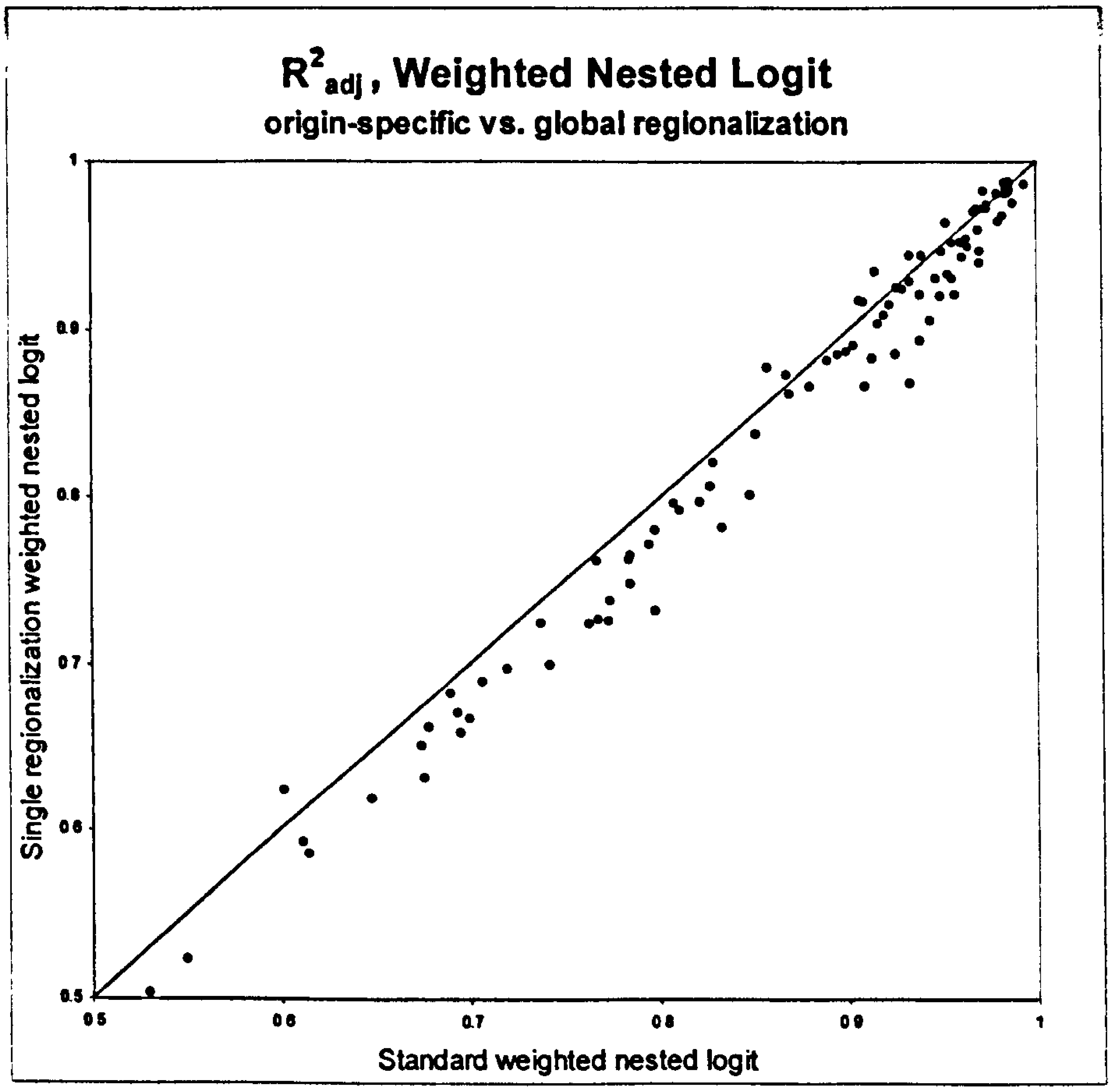


Figure 8.30: Weighted nested logit R^2_{adj} comparing origin-specific & global regionalizations.

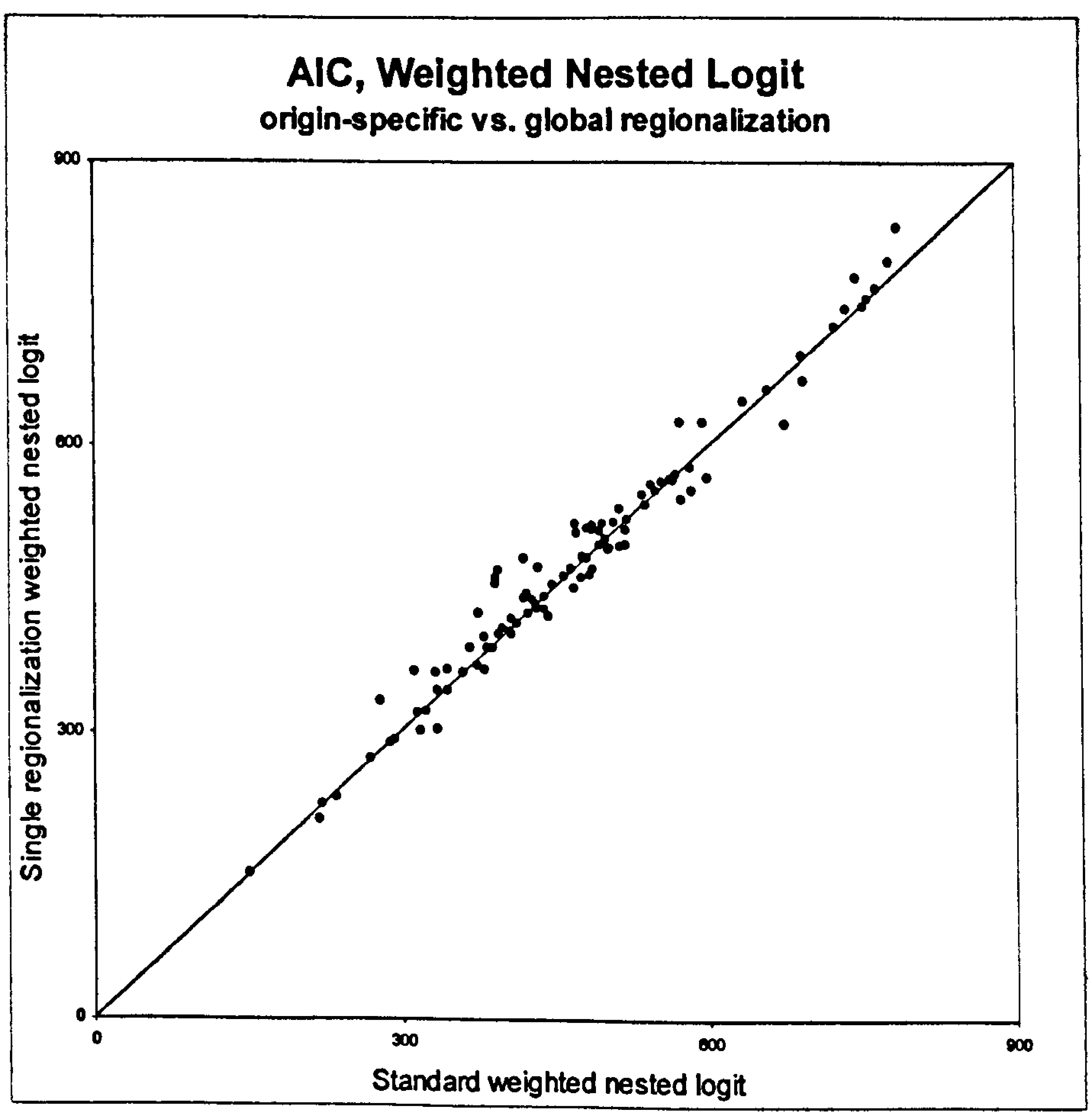


Figure 8.31: Weighted nested logit AIC comparing origin-specific & global regionalizations.

Figures 8.30 and 8.31 shows that there is a weak, but nonetheless noticeable trend for weighted nested logit calibrations made against origin-specific probabilistic regionalizations to provide slightly better goodness-of-fit, on average, than calibrations that use the same regionalization to predict migration from all origins.

The results presented in this section demonstrate that the modification of the discrete nested logit model to utilize a probabilistic approach to regionalization definition successfully reduces sensitivity of the model to the specific regionalization that is used in its calibration. The discrete nested logit model can exhibit noticeable sensitivity to the specific discrete regionalization used in its calibration if it is a region for which the discrete nested logit model's regional utility variable is a significant contributor to the migration destination choice selection. For many origins (just under half, 46 of 100) this does not appear to be a significant factor. This could mean that migrants from those 46% of origins are inherently different from other migrants in that they do not employ hierarchical processing to the same degree, but it is more intuitive to believe that the apparently non-hierarchical (or at least less-hierarchical) behaviour of those 46% of migrants is the results of their discrete regionalizations of space being inappropriate, such that they do not reflect migrants mental maps of destinations sufficiently accurately to be worthwhile, and to positively impact model accuracy.

The author argues that the probabilistic approach to region definition, through aggregation of a large number of semi-randomly generated discrete regionalizations, offers a better approximation to the real world situation where migration from a particular origin is based upon a large number of individuals' different mental maps of space, and their differing information about the various potential destinations.

Of course, it is also debatable whether any individual migrant actually uses a clearly defined discrete regionalization at all. This discrete regionalization is really just a methodological restriction imposed by the traditional method of calibration of the discrete nested logit model. But if we accept for the moment that individuals do create and employ discrete regionalizations of space in their migration decision-making, it is highly unlikely that any two migrants would create the same cognitive regionalization, given the huge number of factors that go into its creation, such as: personal residential and travel experience, geographical spread of extended family and friends...etc. It is obviously much more unlikely that all migrants from a particular origin would somehow arrive at the same discrete view of space. Thus, the author proposes that the method of calibrating the weighted nested logit model presented here is a more appropriate manner in which to apply the nested logit model to the investigation of migration destination choice.

Summary

All variations of the nested logit model have been shown to provide more accurate predictions of migrants' observed behaviour than that produced by the traditional flat-processing migration model. The discrete nested logit model exhibits only marginally better goodness-of-fit for just a small subset of migrant origins. The weighted nested logit model, however, shows more substantial improvements and for many more origins. However, it was a hybrid model combining the benefits of the competing destinations and weighted nested logit model that provided the best overall predictive performance.

It should be noted that examination of the parameter estimates produced by the discrete and weighted nested logit models showed less evidence of multicollinearity than appeared to be evident in the competing destinations model. Because the hybrid weighted nested logit model contains the accessibility variable from the competing destinations model, its parameter estimates also exhibits skew, when compared with traditional model parameter estimates, due to the correlation between that variable and the model's other explanatory variables. However, the improved predictive ability of the hybrid nested logit model appear to justify the inclusion of this variables, suggesting that its benefits outweigh any multicollinearity that it introduces to the model.

All these nested logit models are based upon principles of destination clustering and representing utility at multiple levels of a spatial hierarchy. Indeed the best performing model includes variables to capture two aspects of hierarchical destination choice – accessibility (the likelihood of a destination being cognized within a larger cluster of destinations) and regional utility. Thus the author proposes that these results provide further evidence in support of the theory that internal migration destination choice within Great Britain is an inherently hierarchical process.

Chapter Nine

Comparing Hierarchical Models of Destination Choice

The two preceding chapters have focused on comparing the predictive abilities of hierarchical models with the traditional, flat-processing model of migration destination choice. These comparisons have demonstrated that both the competing destinations and nested logit family of hierarchical models predict migrants' destination choice more accurately than the traditional model for some origins, with little or no change in the accuracy of predictions for other origins. It is useful now to compare the similarities and differences of hierarchical approaches directly.

In this chapter results from the discrete and weighted nested logit models are compared in order to demonstrate that the weighted model offers generally superior performance to the discrete model. Once it has been demonstrated to be the 'better' of the two nested logit models, results from the weighted nested logit model are then compared with those from the competing destinations model, highlighting any spatial variation in how these two models predict migration and exploring differences in their predicted explanatory variables' parameter estimates.

Sufficient differences were evident from this comparison of the competing destinations and weighted nested logit approaches to suggest that the derivation and application of a hybrid model combining the benefits of both models might prove worthwhile.

The results of calibrating this hybrid weighted nested logit model are compared with those from the competing destinations and weighted nested logit models in order to demonstrate that the hybrid model does indeed offer worthwhile improvements over the more basic hierarchical models.

The goodness-of-fit of the various models is examined first, followed by a consideration of the parameter estimates predicted by each model. Particular attention is paid to the spatial variation in goodness-of-fit and parameter estimates between the models, as this is particularly useful when determining whether they are capturing the same or different aspects of migrants' hierarchical decision-making processes.

Goodness-of-Fit

As in chapters 7 and 8, goodness-of-fit is assessed and compared by three methods: plotting R^2_{adj} and AIC values, mapping spatial patterns in R^2_{adj} and AIC values, and, mapping the residual flows between the two models' predictions.

Discrete and Weighted Nested Logit Models

The predictive ability of discrete and weighted nested logit models was directly compared by scatter-plotting the R^2_{adj} and AIC values for 100 pairs of origin-specific model calibrations, see figures 9.1 and 9.2 below.

It is clear from figures 9.1 and 9.2 that there is a general trend for the weighted nested logit model to provide a goodness-of-fit improvement (i.e. higher R^2_{adj} and lower AIC values) over the discrete nested logit model, for the majority of origins.

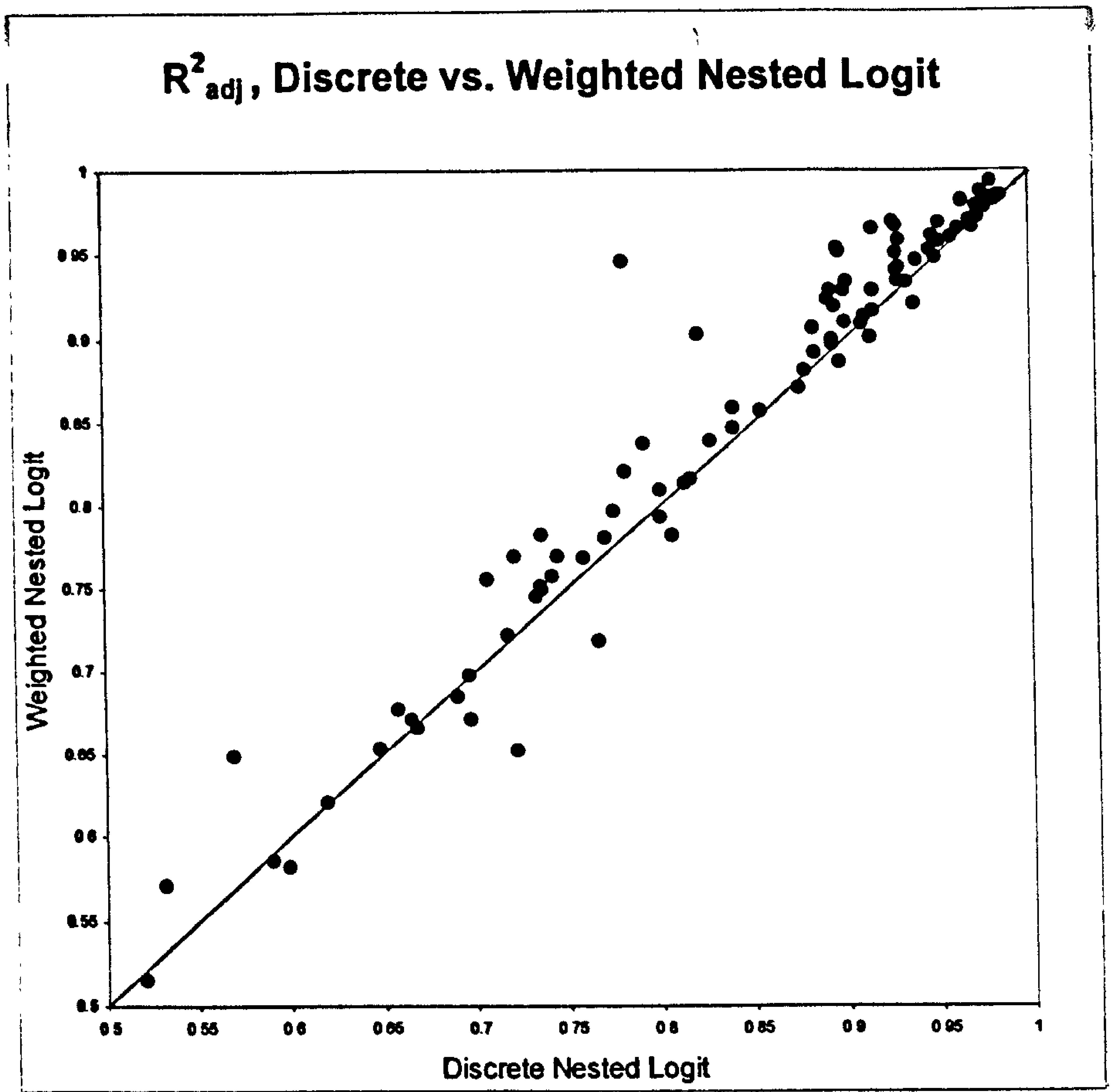


Figure 9.1: R^2_{adj}: discrete vs. weighted nested logit models

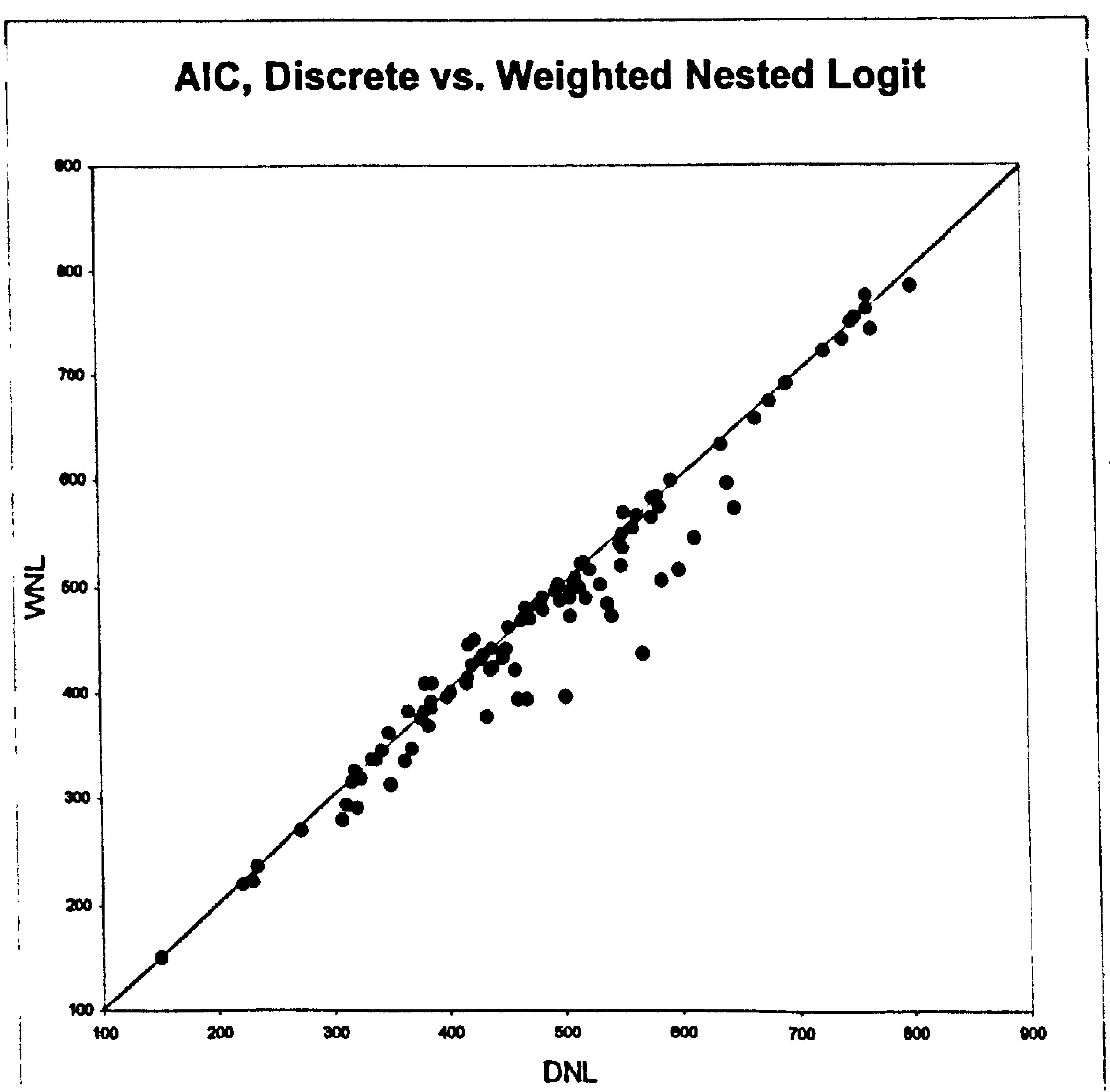
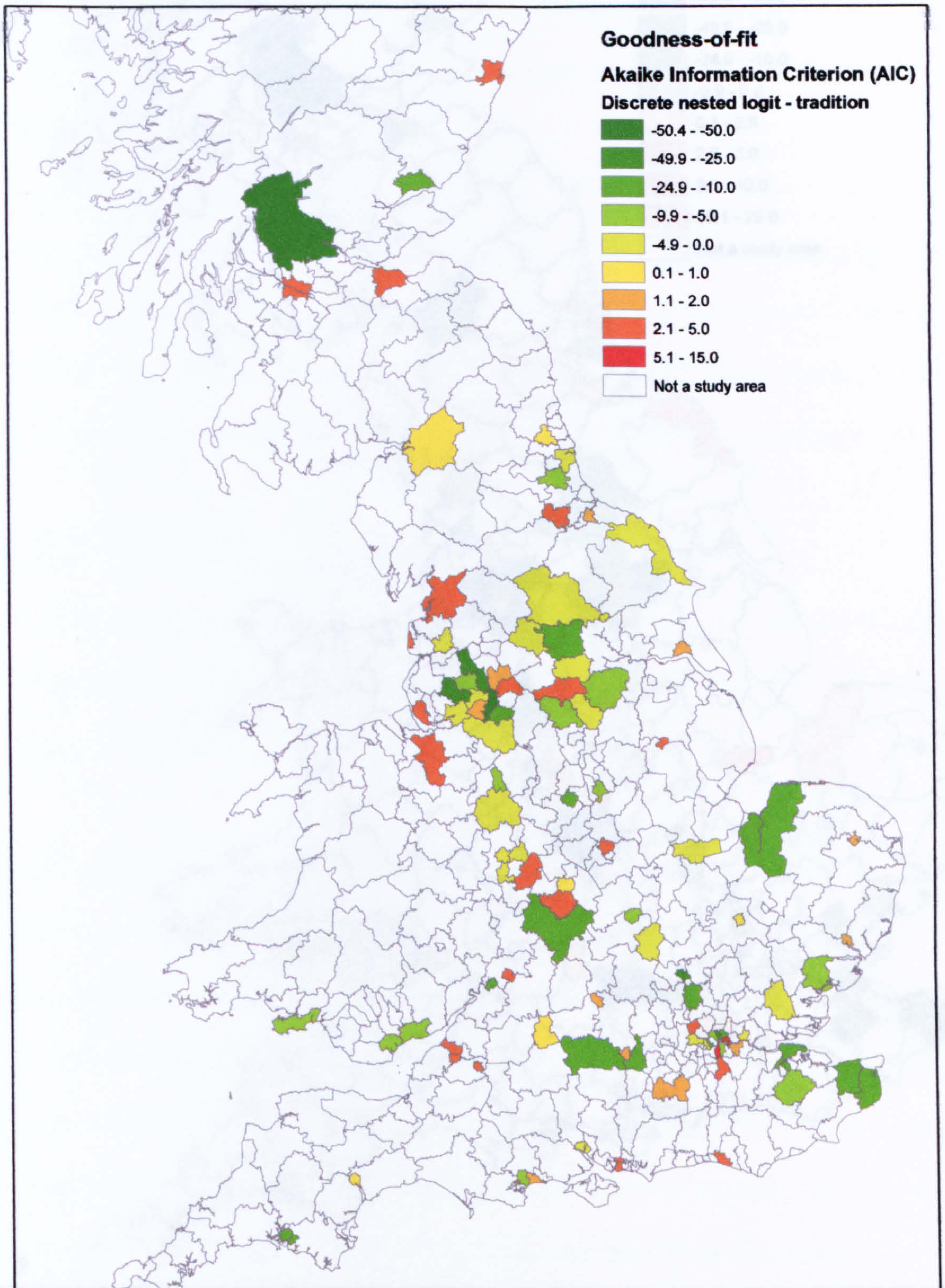
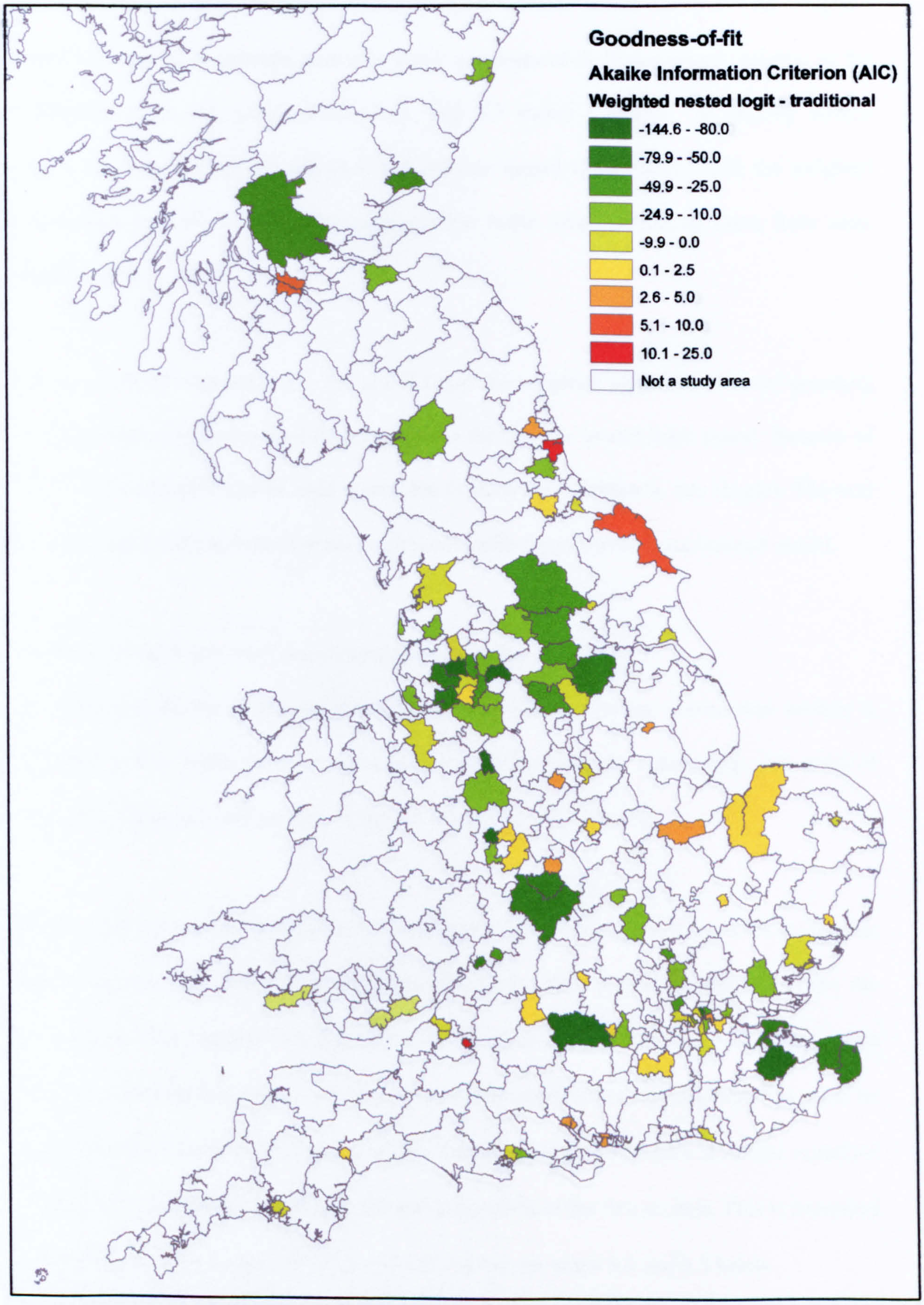


Figure 9.2: AIC: discrete vs. weighted nested logit models

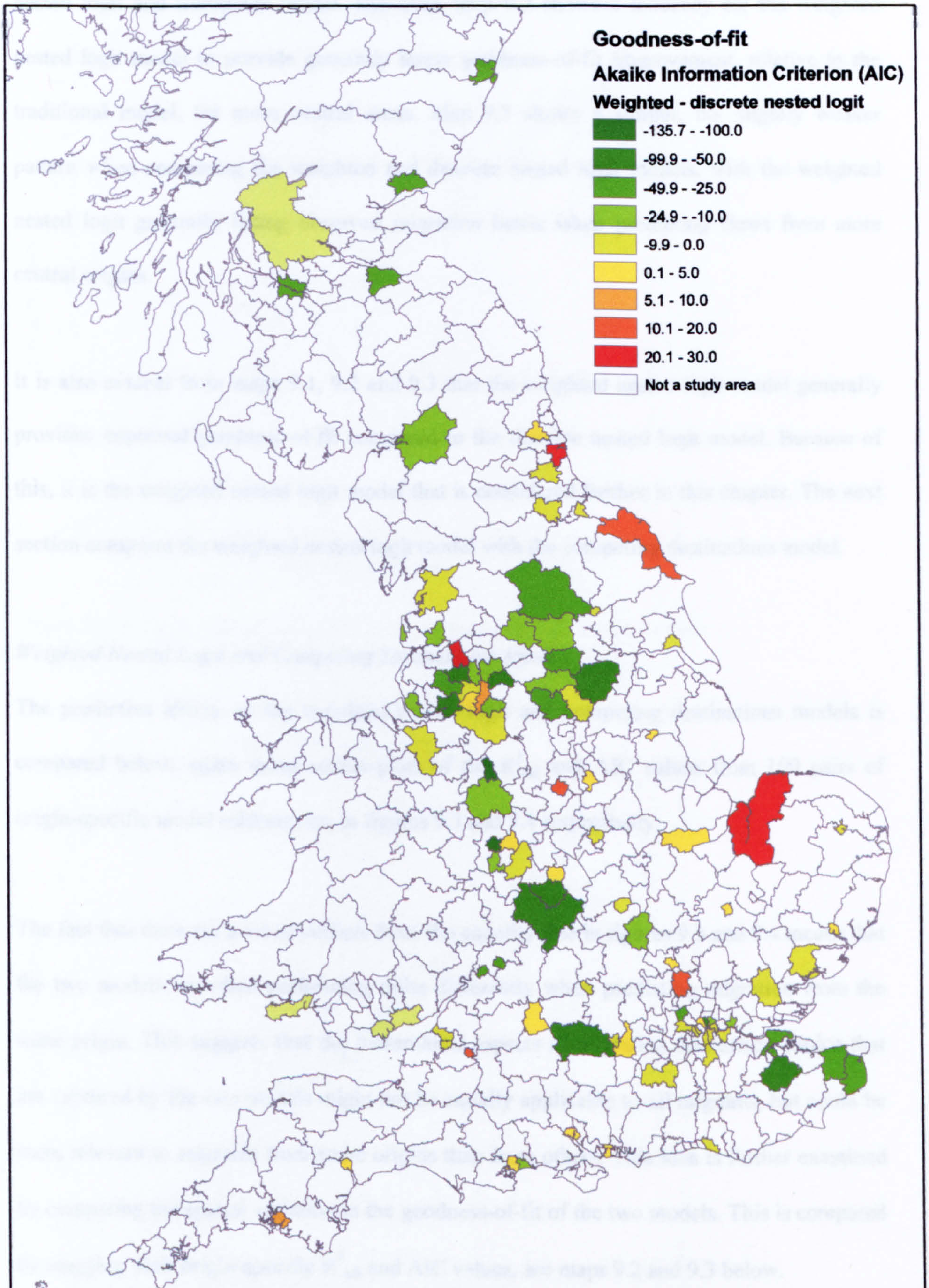
In order to highlight any differences in the spatial patterns of goodness-of-fit of the two models, maps 9.1 and 9.2, below, present the spatial variation in the goodness-of-fit of the discrete and weighted nested logit models, respectively, compared to the traditional model, and then map 9.3 compares their goodness-of-fit directly.



Map 9.1: Change in AIC: discrete nested logit – traditional models.



Map 9.2: Change in AIC: weighted nested logit – traditional models.



Map 9.3: Change in AIC: weighted – discrete nested logit models.

Map 9.1 shows no major difference in the spatial patterns of goodness-of-fit of the discrete nested logit and traditional model. However, map 9.2 shows a tendency for the weighted nested logit model to provide generally larger goodness-of-fit improvement, relative to the traditional model, for more central areas. Map 9.3 shows a similar, but slightly weaker pattern when comparing the weighted and discrete nested logit models, with the weighted nested logit generally fitting observed migration better when predicting flows from more central origins.

It is also evident from maps 9.1, 9.2 and 9.3 that the weighted nested logit model generally provides improved goodness-of-fit compared to the discrete nested logit model. Because of this, it is the weighted nested logit model that is considered further in this chapter. The next section compares the weighted nested logit model with the competing destinations model.

Weighted Nested Logit and Competing Destinations Models

The predictive ability of the weighted nested logit and competing destinations models is compared below, again using scatter-plots of the R^2_{adj} and AIC values from 100 pairs of origin-specific model calibrations, in figures 9.3 and 9.4 respectively.

The fact that there are a lot of outliers from the equality line in figures 9.3 and 9.4 means that the two models are often performing quite differently when predicting migration from the same origin. This suggests that the hierarchical aspects of migration destination choice that are captured by the two models might not be equally applicable to all migrants, but could be more relevant to migrants from some origins than from others. This idea is further examined by comparing the spatial variation in the goodness-of-fit of the two models. This is compared by mapping their origin-specific R^2_{adj} and AIC values, see maps 9.2 and 9.3 below.

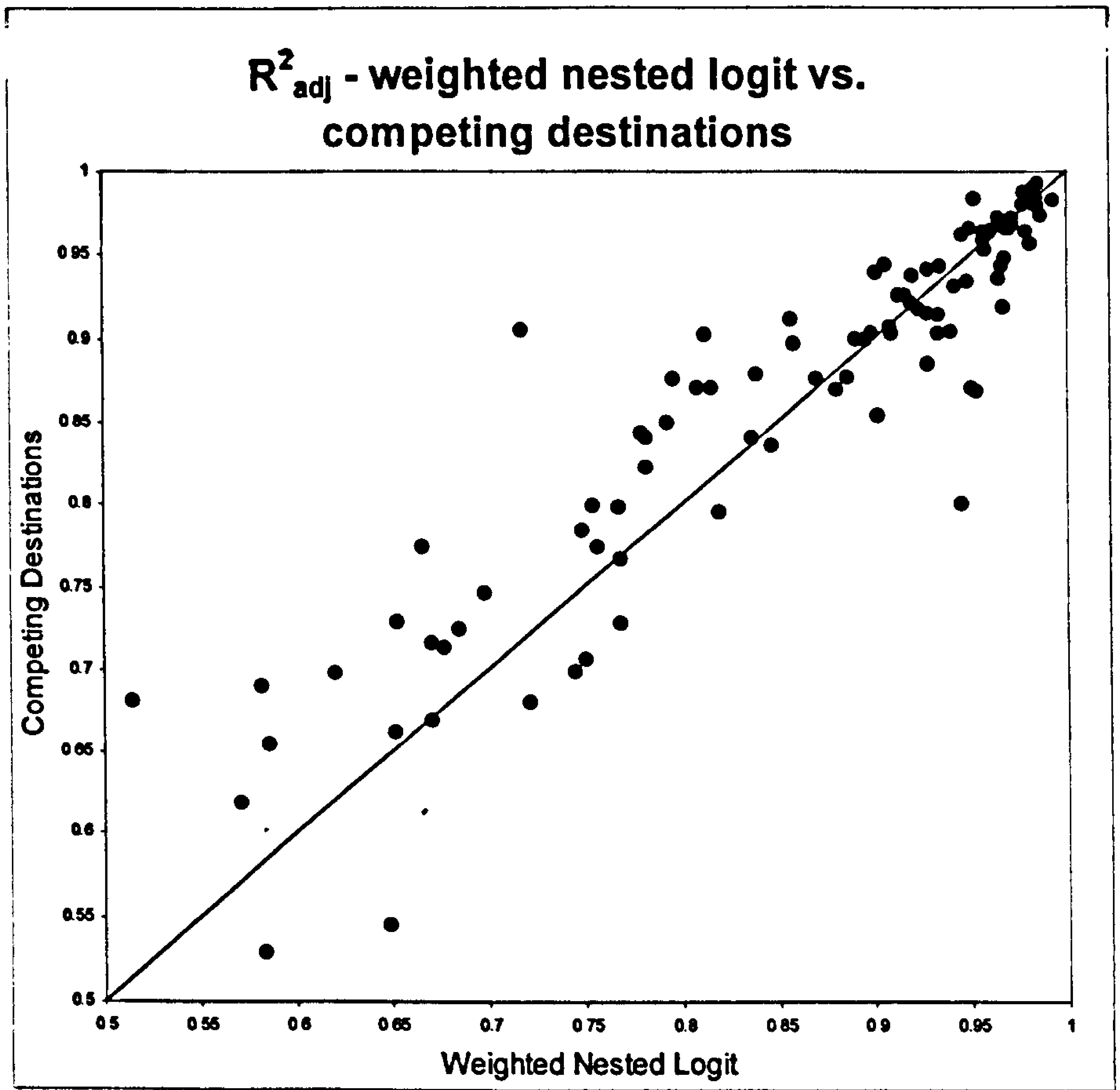


Figure 9.3: R^2_{adj}: weighted nested logit and competing destinations models

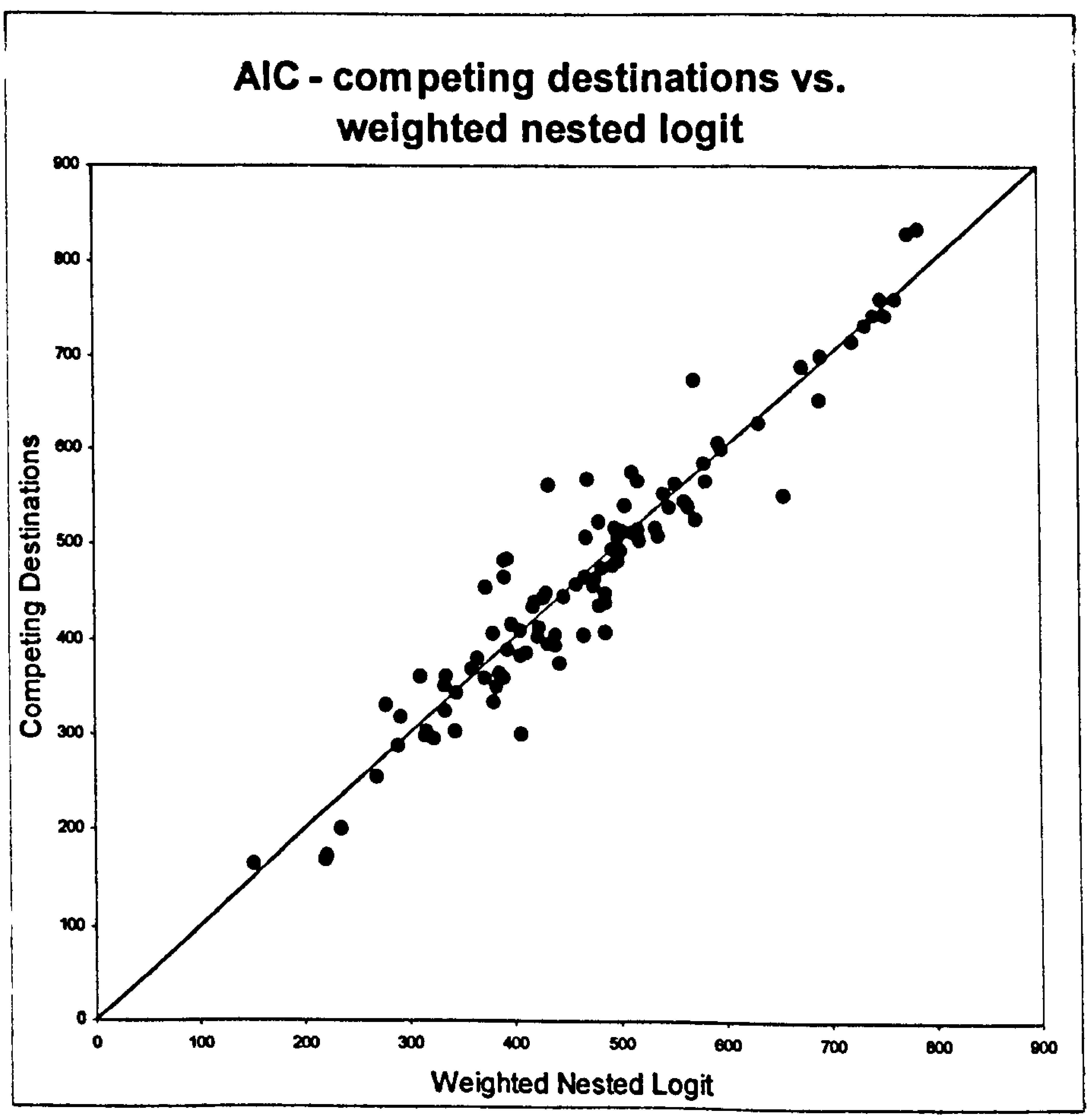
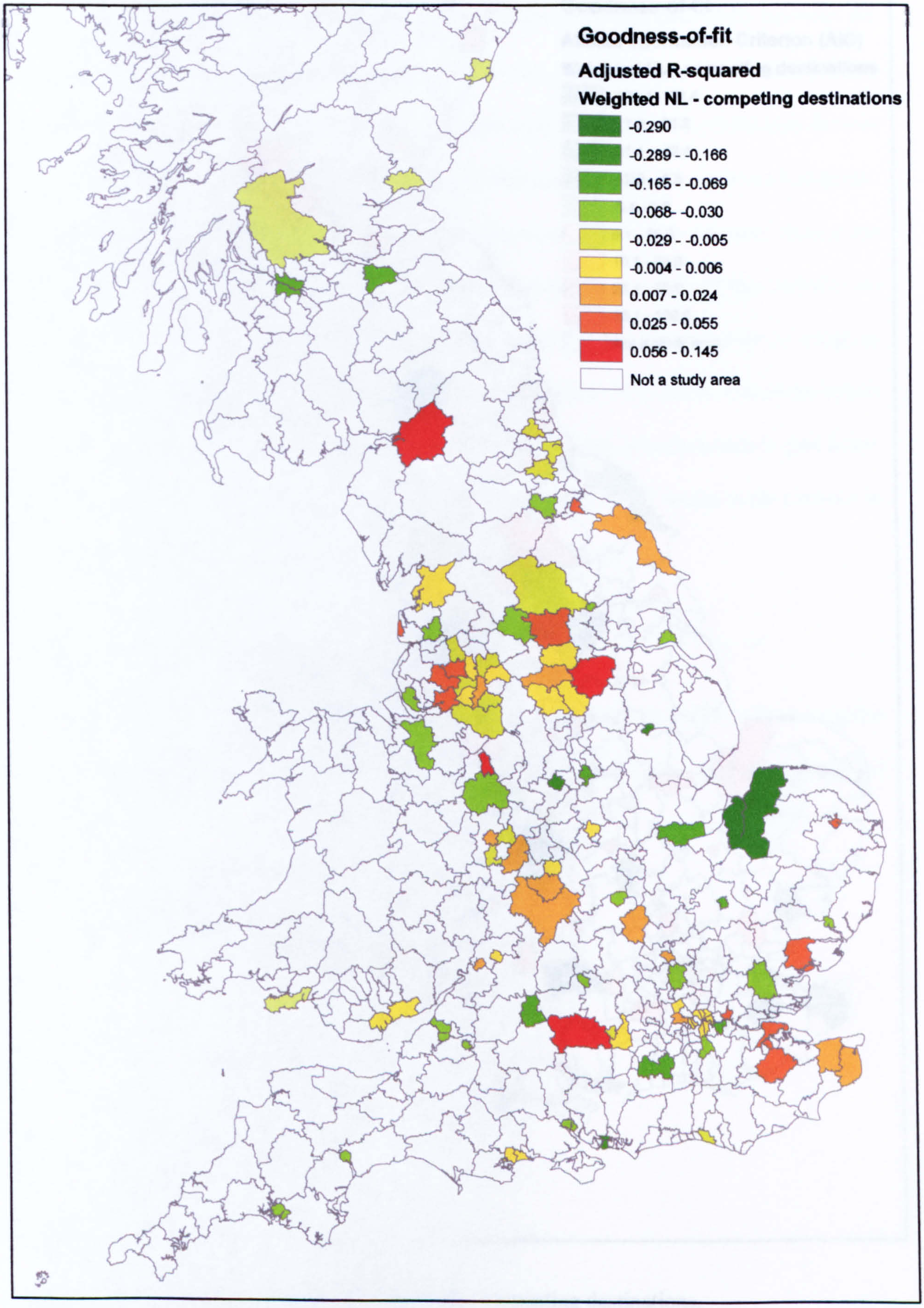
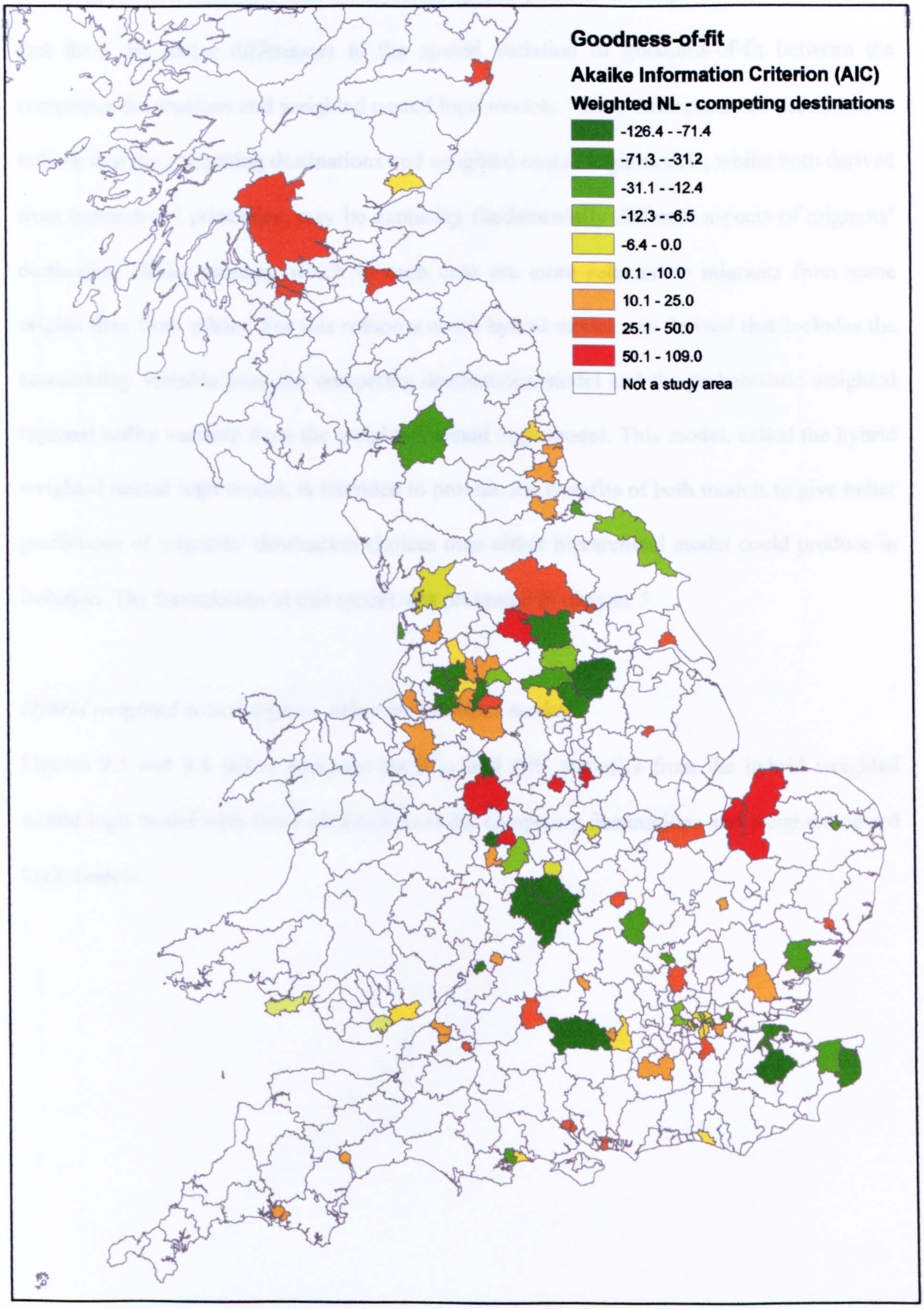


Figure 9.4: AIC: weighted nested logit and competing destinations models



Map 9.4: R^2_{adj} difference, weighted nested logit - competing destinations.



Map 9.5: AIC difference, weighted nested logit - competing destinations.

The extensive variation yet lack of any clear spatial patterns on maps 9.4 and 9.5 indicates that there are major differences in the spatial variation of goodness-of-fit between the competing destinations and weighted nested logit models. These differences led the author to believe that the competing destinations and weighted nested logit models, whilst both derived from hierarchical principles, may be capturing fundamentally different aspects of migrants' destination choice process, which in each case are more relevant to migrants from some origins than from others. For this reason a novel hybrid model was derived that includes the accessibility variable from the competing destinations model and the probabilistic weighted regional utility variable from the weighted nested logit model. This model, called the hybrid weighted nested logit model, is intended to provide the benefits of both models to give better predictions of migrants' destination choices than either hierarchical model could produce in isolation. The formulation of this model was presented in chapter 5.

Hybrid weighted nested logit vs. other hierarchical models

Figures 9.5 and 9.6 below compare the R^2_{adj} and AIC statistics from the hybrid weighted nested logit model with those obtained from the competing destinations and weighted nested logit models.

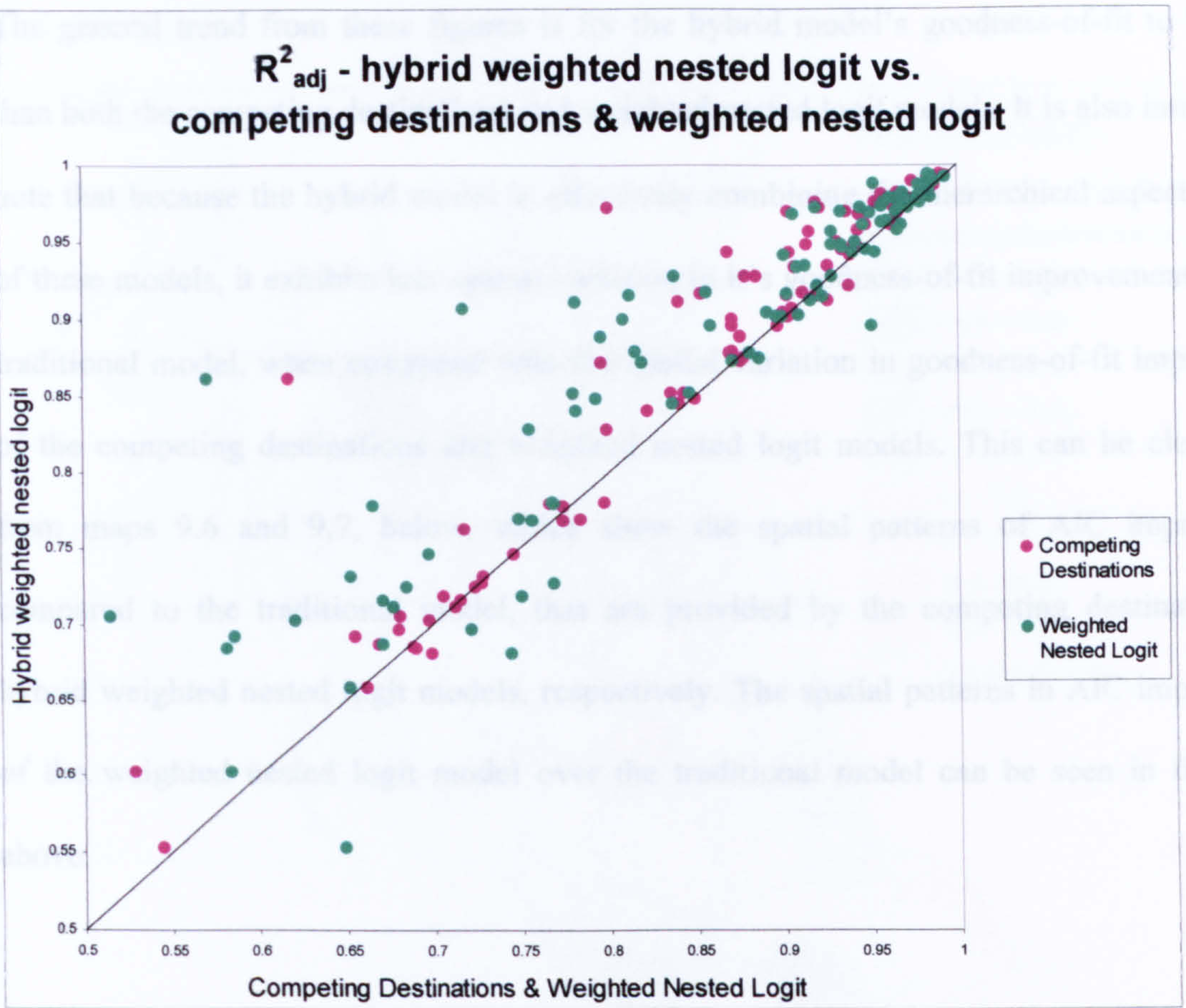


Figure 9.5: R^2_{adj}: hybrid, competing destinations and weighted nested logit models

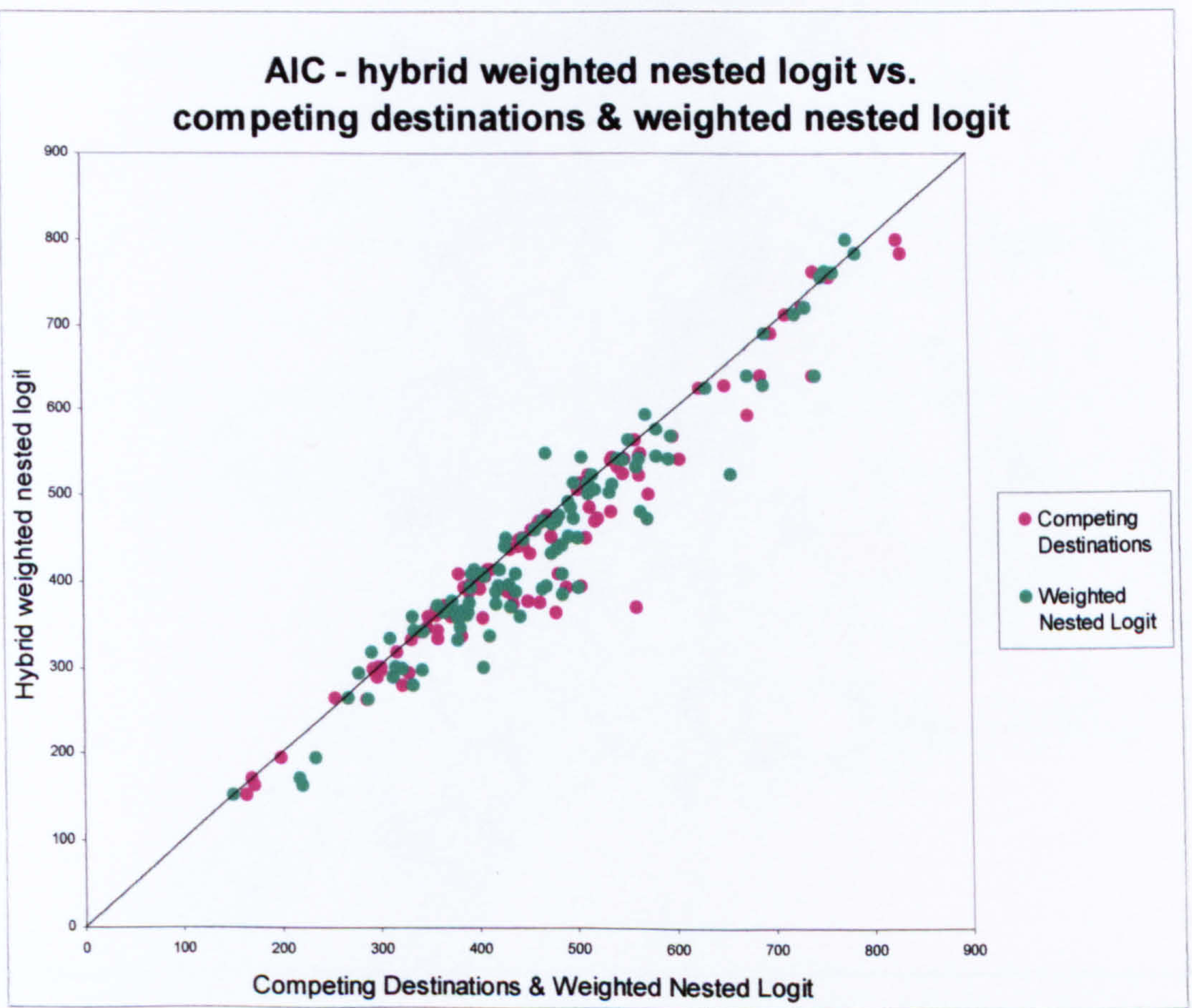
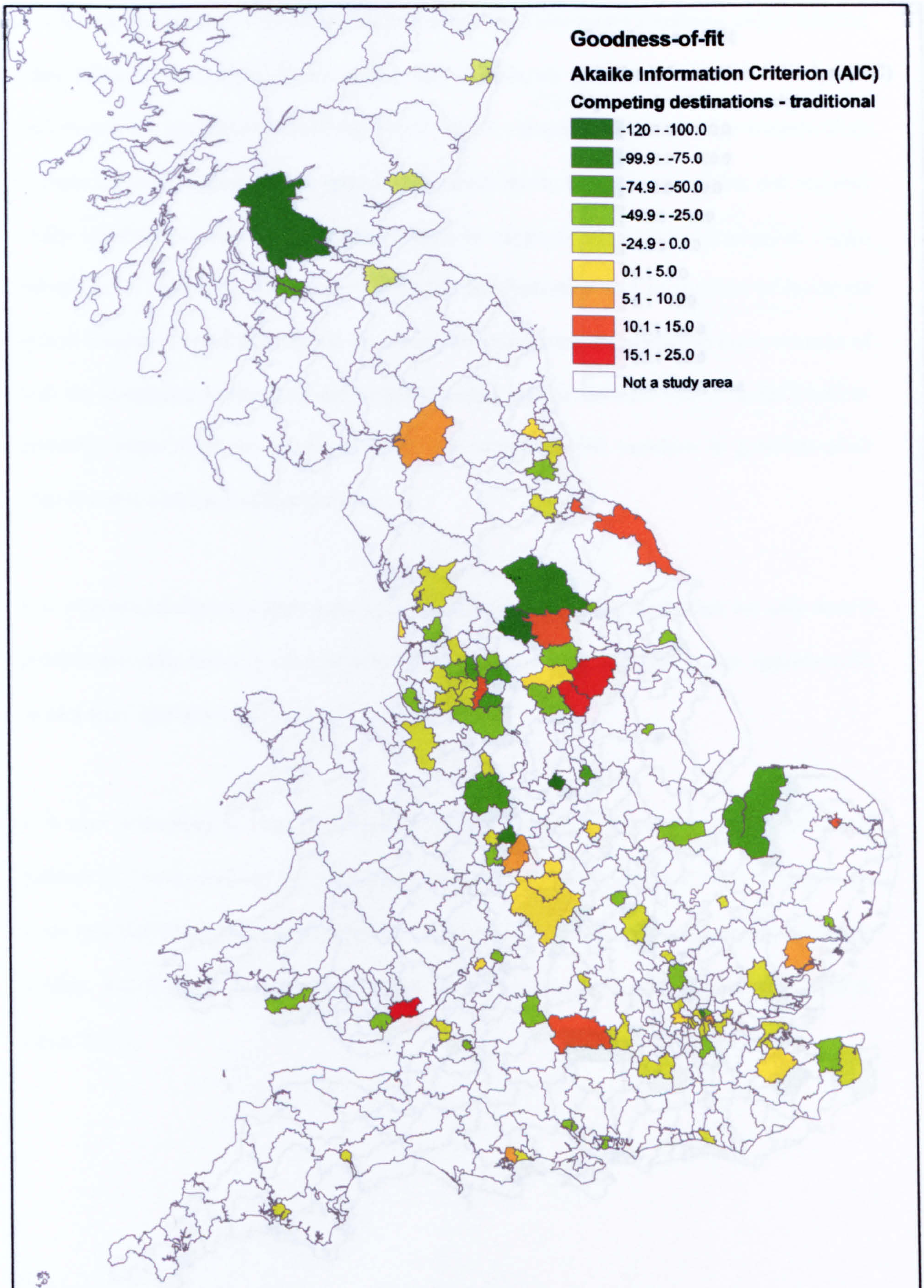
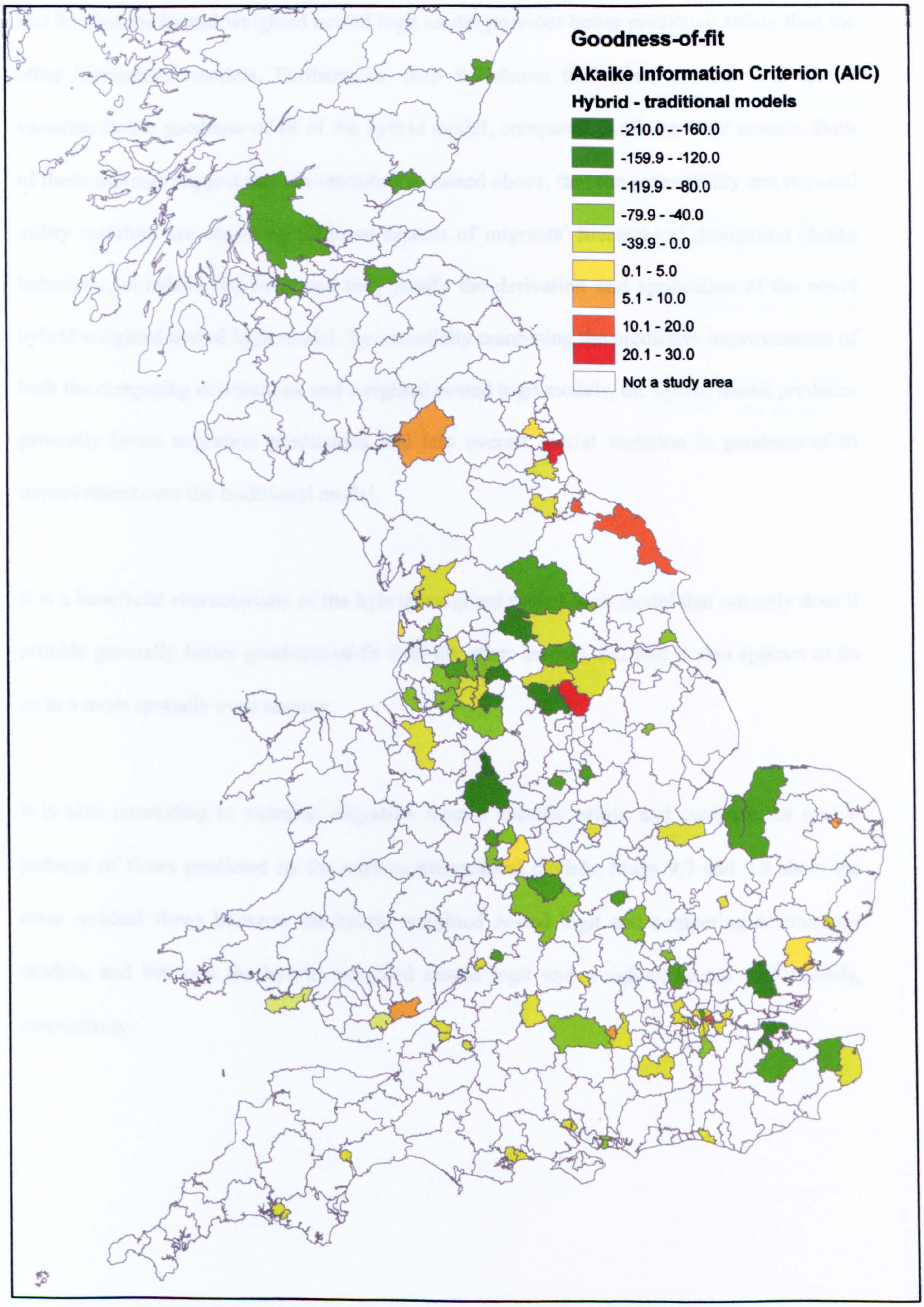


Figure 9.6: AIC: hybrid, competing destinations and weighted nested logit models

The general trend from these figures is for the hybrid model's goodness-of-fit to be better than both the competing destinations and weighted nested logit models. It is also interested to note that because the hybrid model is effectively combining the hierarchical aspects of both of these models, it exhibits less spatial variation in it's goodness-of-fit improvement over the traditional model, when compared with the spatial variation in goodness-of-fit improvement of the competing destinations and weighted nested logit models. This can be clearly seen from maps 9.6 and 9.7, below, which show the spatial patterns of AIC improvement, compared to the traditional model, that are provided by the competing destinations and hybrid weighted nested logit models, respectively. The spatial patterns in AIC improvement of the weighted nested logit model over the traditional model can be seen in figure 9.2, above.



Map 9.6: AIC difference, competing destinations - traditional models.

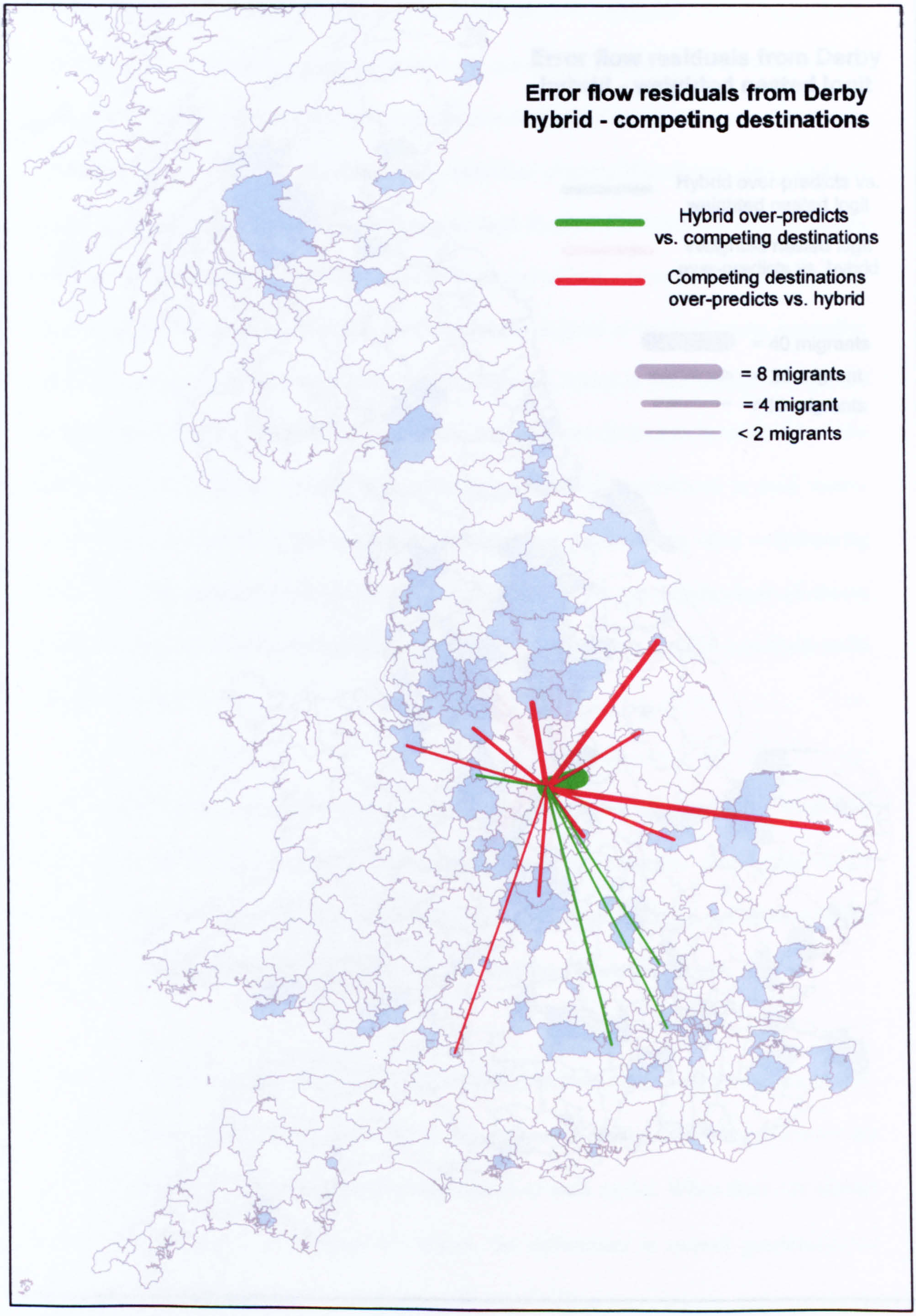


Map 9.7: AIC difference, hybrid weighted nested logit - traditional models.

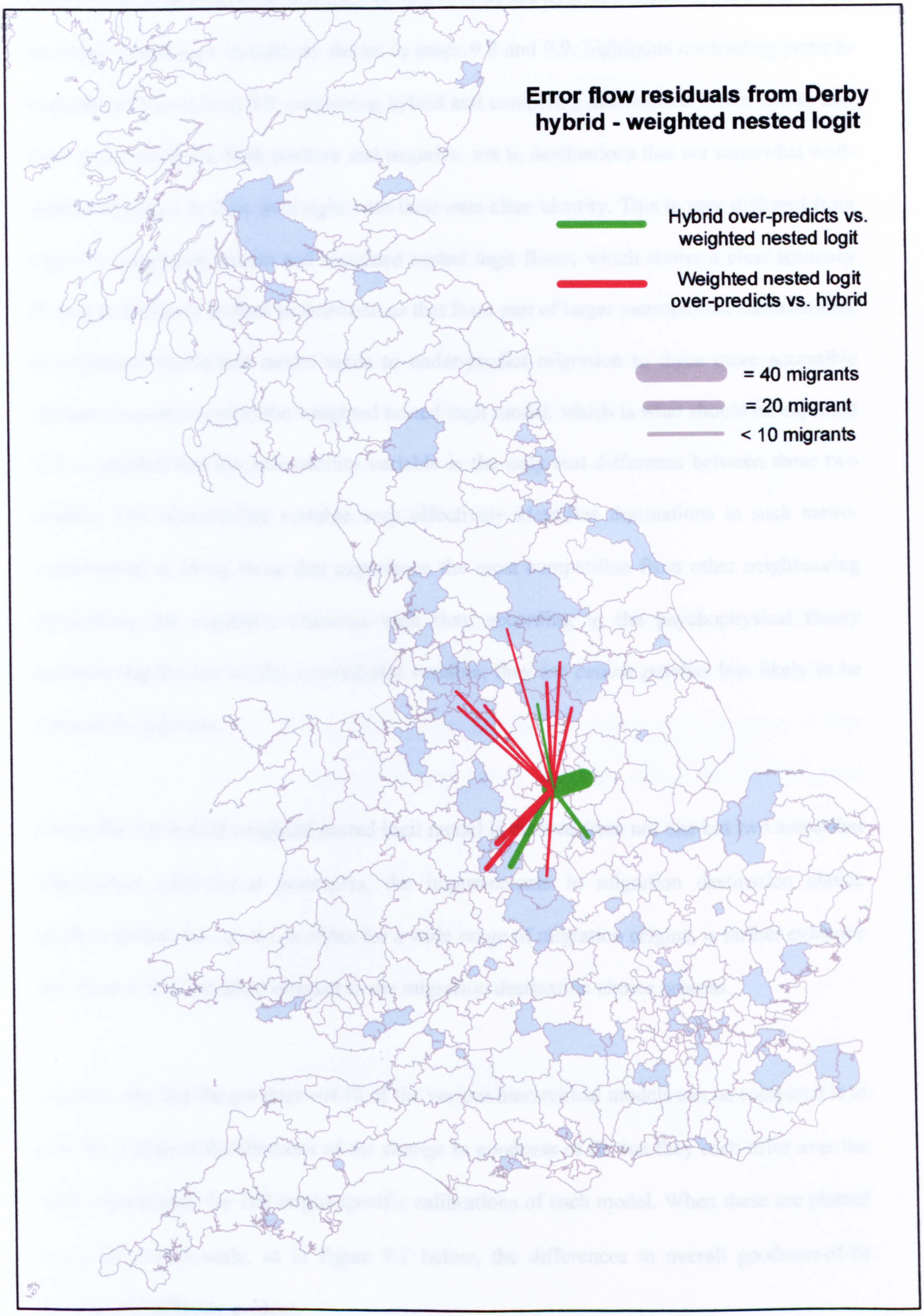
Map 9.7, when compared with maps 9.2 and 9.6, confirms the conclusion from figures 9.5 and 9.6 that the hybrid weighted nested logit model provides better predictive ability than the other hierarchical models. Furthermore, map 9.7 shows that there is much less spatial variation in the goodness-of-fit of the hybrid model, compared to all the other models. Both of these findings suggest that the speculation, raised above, that the accessibility and regional utility variables are capturing different aspects of migrants' hierarchical destination choice behaviour, is indeed the case, and they justify the derivation and application of the novel hybrid weighted nested logit model. By essentially combining the predictive improvements of both the competing destinations and weighted nested logit models, the hybrid model produces generally better migration predictions and less overall spatial variation in goodness-of-fit improvement over the traditional model.

It is a beneficial characteristic of the hybrid weighted nested logit model that not only does it provide generally better goodness-of-fit than the other models, but that it also appears to do so in a more spatially even manner.

It is also interesting to examine migration from a specific origin and compare the spatial patterns of flows predicted by the various hierarchical models. Maps 9.7 and 9.8 show the error residual flows between the hybrid weighted nested logit and competing destinations models, and between the hybrid weighted nested logit and weighted nested logit models, respectively.



Map 9.8: Residual error flows from Derby, hybrid - competing destinations models.



Map 9.9: Residual error flows from Derby, hybrid – weighted nested logit models.

Comparison of the migration predicted from Derby by the hybrid, competing destinations and weighted nested logit models, as shown in maps 9.8 and 9.9, highlights contrasting patterns of predicted flows. Map 9.8, comparing hybrid and competing destinations flows, shows that most major residuals, both positive and negative, are to destinations that are somewhat well-defined as places in their own right with their own clear identity. This is very different from map 9.9, comparing hybrid and weighted nested logit flows, which shows a clear tendency for outlier residuals to flow to destinations that form part of larger metropolitan conurbations. In particular, the hybrid model tends to under-predict migration to these more accessible destinations, compared to the weighted nested logit model, which is what should be expected if it is recalled that the accessibility variable is the only real difference between these two models. This accessibility variable very effectively identifies destinations in such metro-conurbations as being those that experience the most competition from other neighbouring destinations for migrant's attention such that, according to the psychophysical theory underpinning the use of the accessibility variable, they are *ceteris paribus* less likely to be selected by migrants.

Given that the hybrid weighted nested logit model is derived from not one but two somewhat independent hierarchical principles, the improvements in migration destination choice predictions that this model provides for a wide range of migration origins, is further evidence that there is a hierarchical element to the migration destination choice process.

Another way that the goodness-of-fit of the various hierarchical models can be compared is to plot the statistical distributions of the change in goodness of fit that they each offer over the traditional model, for 100 origin-specific calibrations of each model. When these are plotted using the y-axis scale, as in figure 9.7 below, the differences in overall goodness-of-fit become immediately evident.

In figure 9.7 the hierarchical models have been ranked top to bottom in improving goodness-of-fit order: discrete nested logit, weighted nested logit model, competing destinations model and finally the hybrid weighted nested logit model.

Examining the statistical distributions of the models' goodness-of-fit improvements over the traditional model highlights that not only does the hybrid weighted nested logit model provide the largest goodness-of-fit improvements, but also, the hybrid model improves goodness-of-fit for more of the 100 selected migration origins than does any other model. Though there do remain a few origins for which the extra complexity of the hybrid model is not sufficiently offset by its better migration flow predictions, such that hybrid model AIC values are slightly higher than from the tradition model.

In summary, the above comparison of goodness-of-fit of the competing destinations and weighted nested logit models has appeared to justify use of a hybrid approach that combines the hierarchical benefits of both models. Presentation and discussion of the accuracy of flow predictions from such a hybrid weighted nested logit model has shown that it does indeed provide the best goodness-of-fit to observed migration behaviour, of all the models applied here.

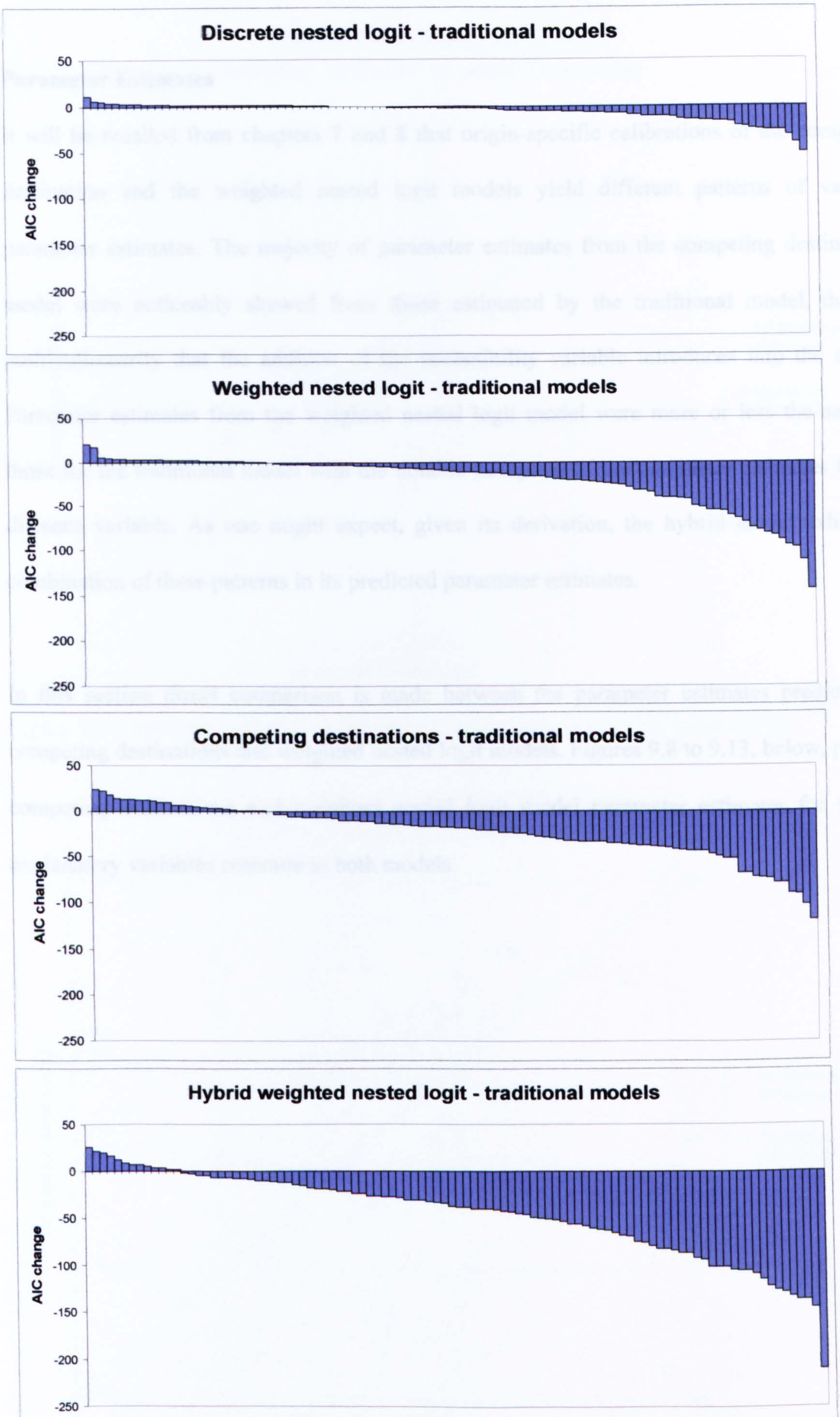


Figure 9.7. Distributions of AIC change, hierarchical model vs. traditional.

Parameter Estimates

It will be recalled from chapters 7 and 8 that origin-specific calibrations of the competing destination and the weighted nested logit models yield different patterns of variable parameter estimates. The majority of parameter estimates from the competing destinations model were noticeably skewed from those estimated by the traditional model, due the multicollinearity that the addition of the accessibility variable introduces into the model. Parameter estimates from the weighted nested logit model were more or less the same as those for the traditional model with the notable exception of the parameter estimates for the distance variable. As one might expect, given its derivation, the hybrid model exhibits a combination of these patterns in its predicted parameter estimates.

In this section direct comparison is made between the parameter estimates predicted by competing destinations and weighted nested logit models. Figures 9.8 to 9.13, below, plot the competing destinations and weighted nested logit model parameter estimates for the six explanatory variables common to both models.

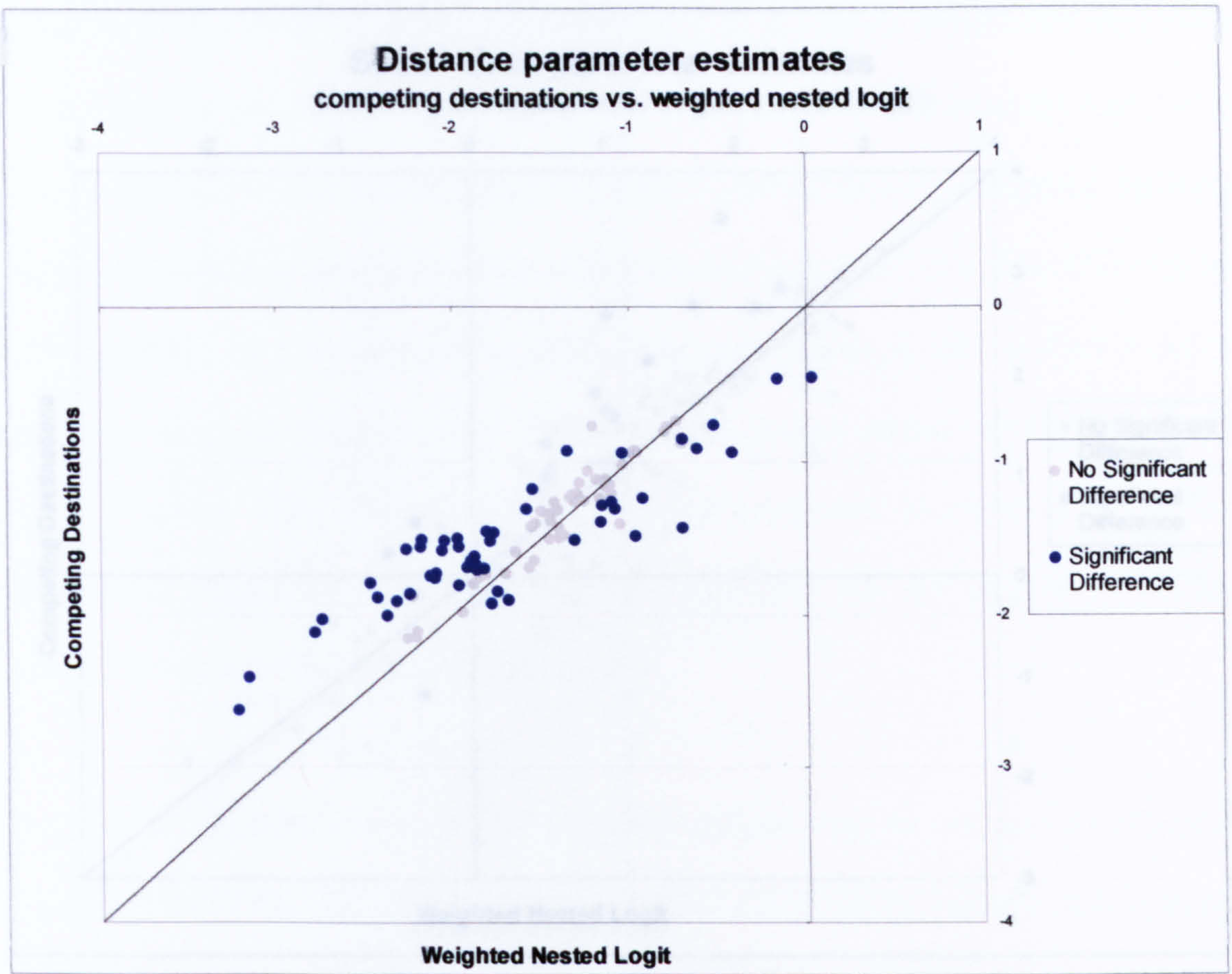


Figure 9.8: Distance parameter estimates: competing destinations vs. weighted NL

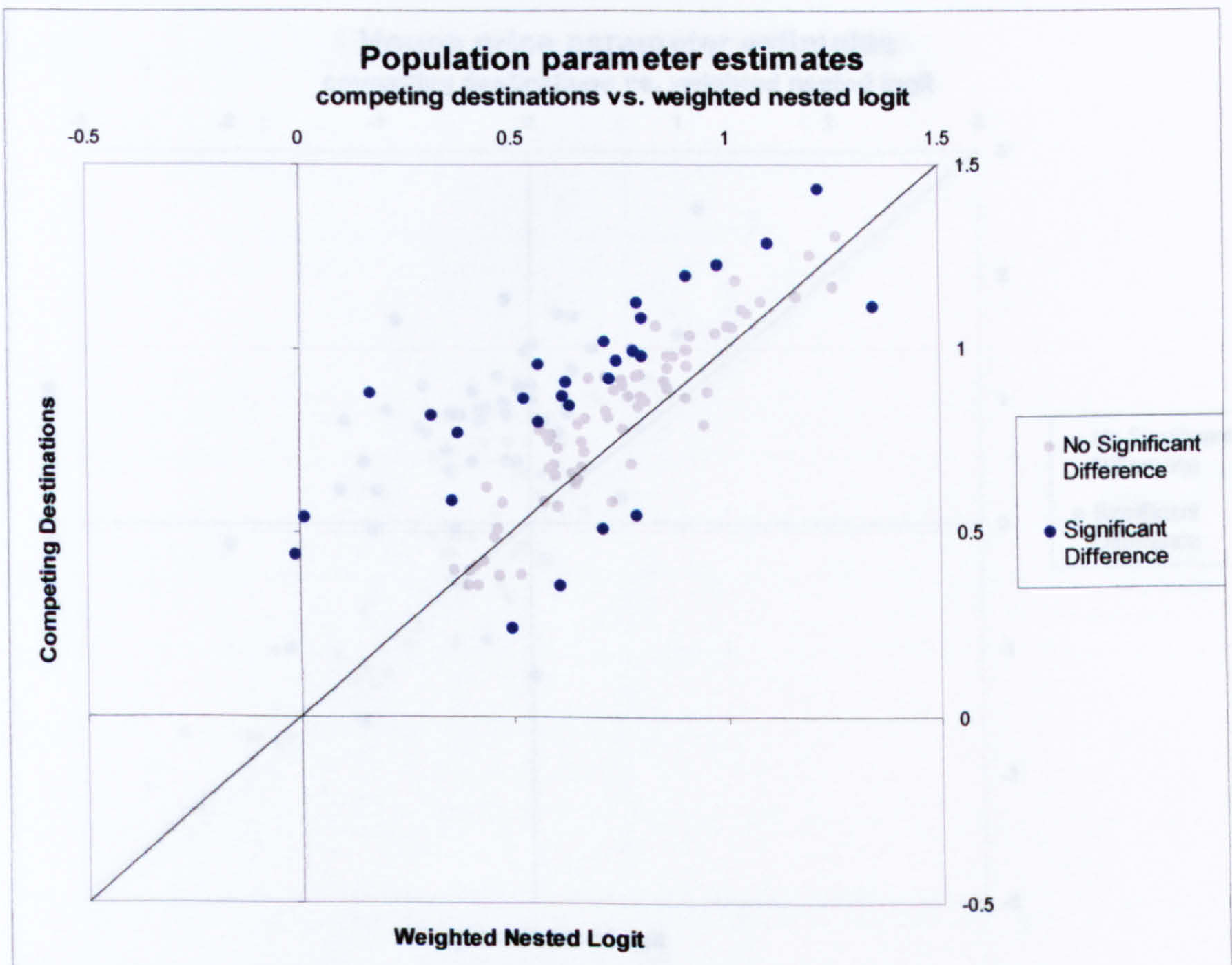


Figure 9.9: Population parameter estimates: competing destinations vs. weighted NL

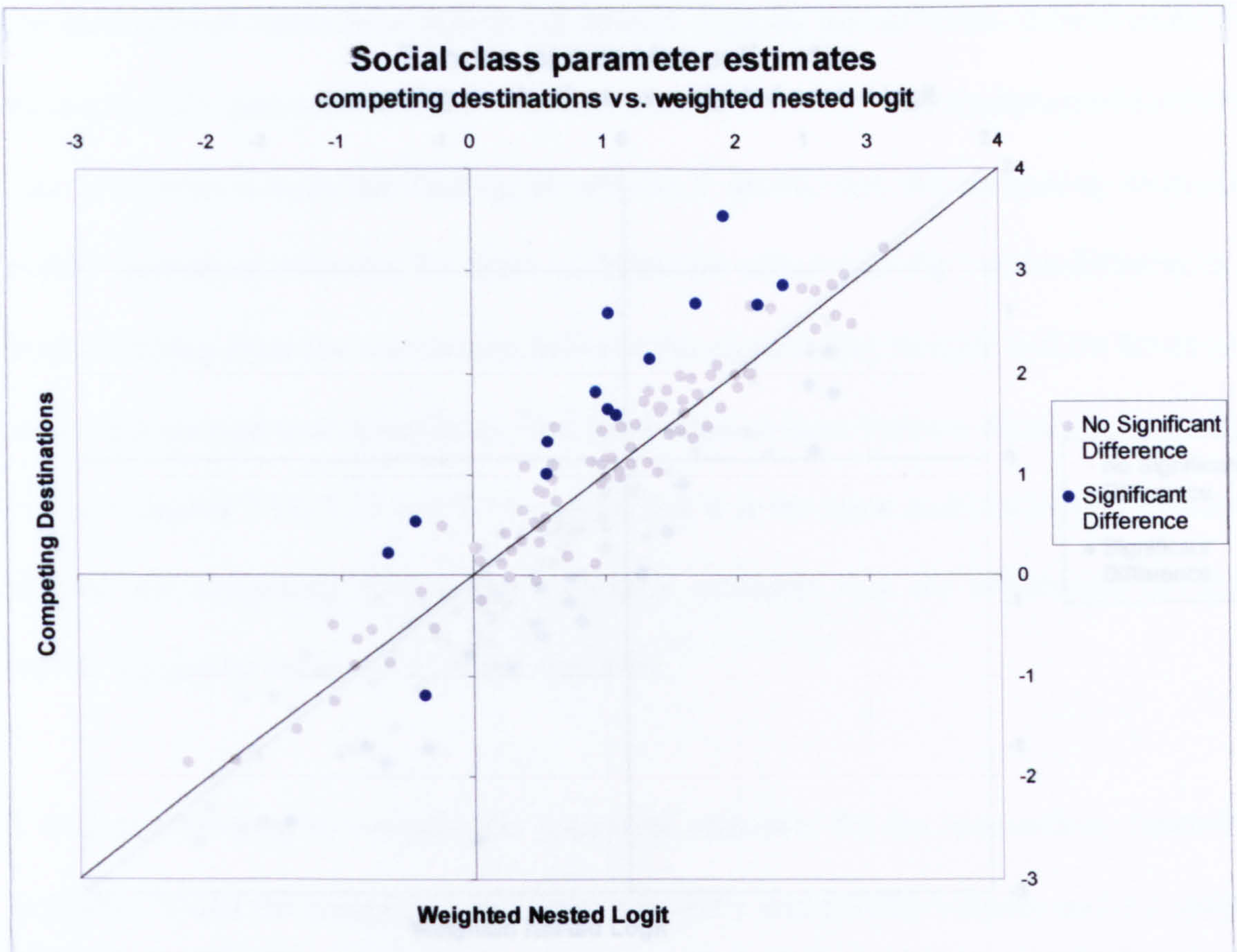


Figure 9.10: Social class parameter estimates: competing destinations vs. weighted NL

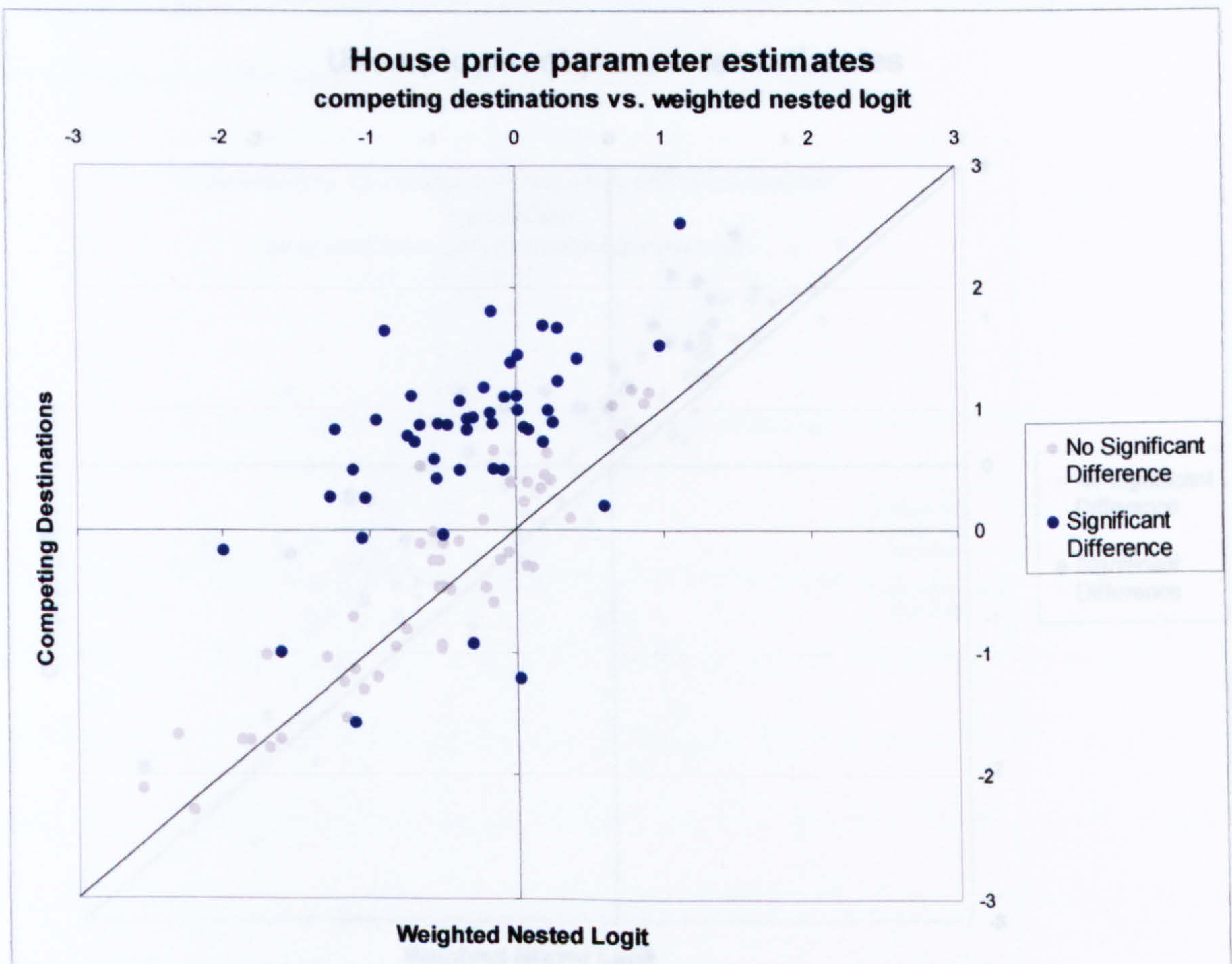


Figure 9.11: House price parameter estimates: competing destinations vs. weighted NL

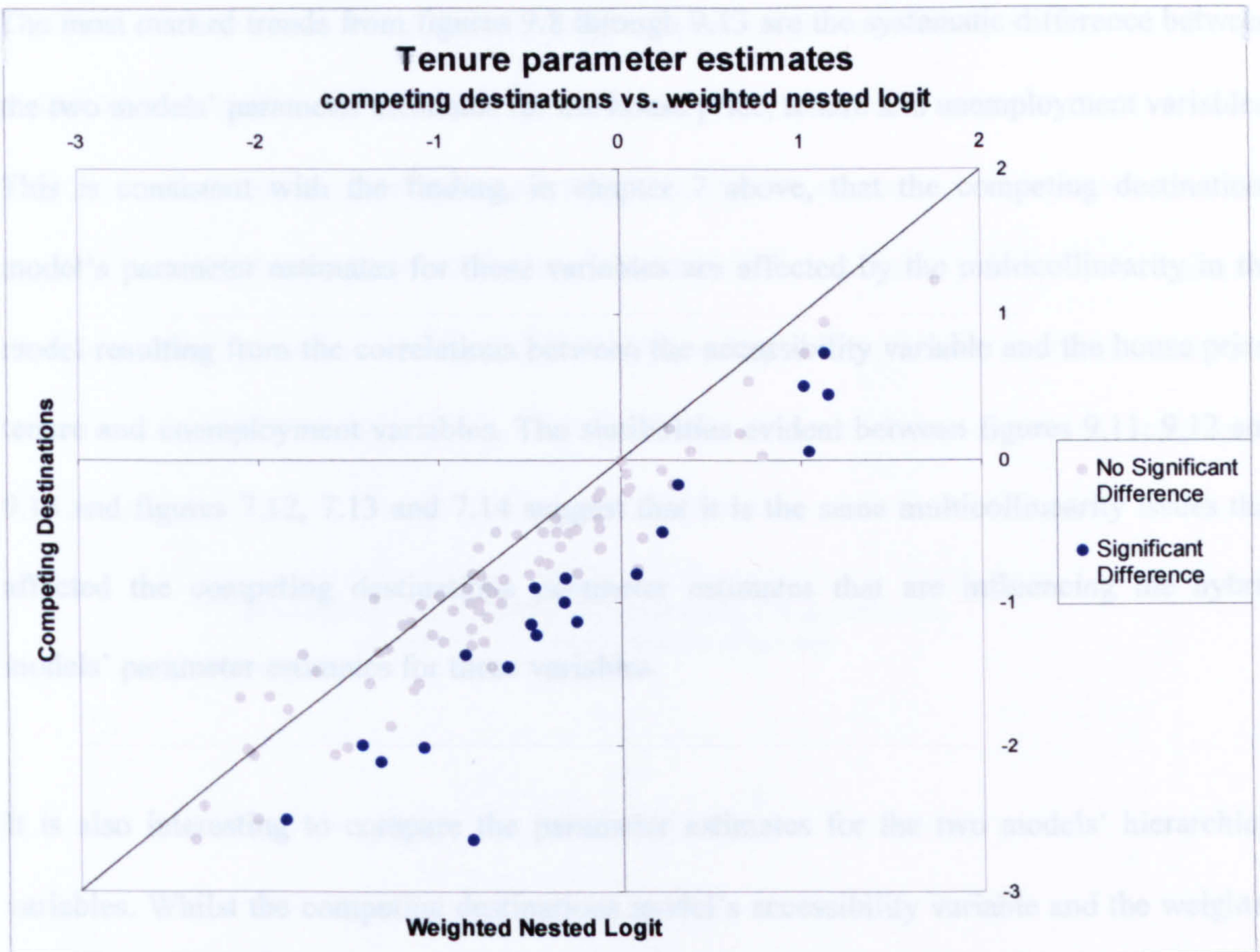


Figure 9.12: Tenure parameter estimates: competing destinations vs. weighted NL

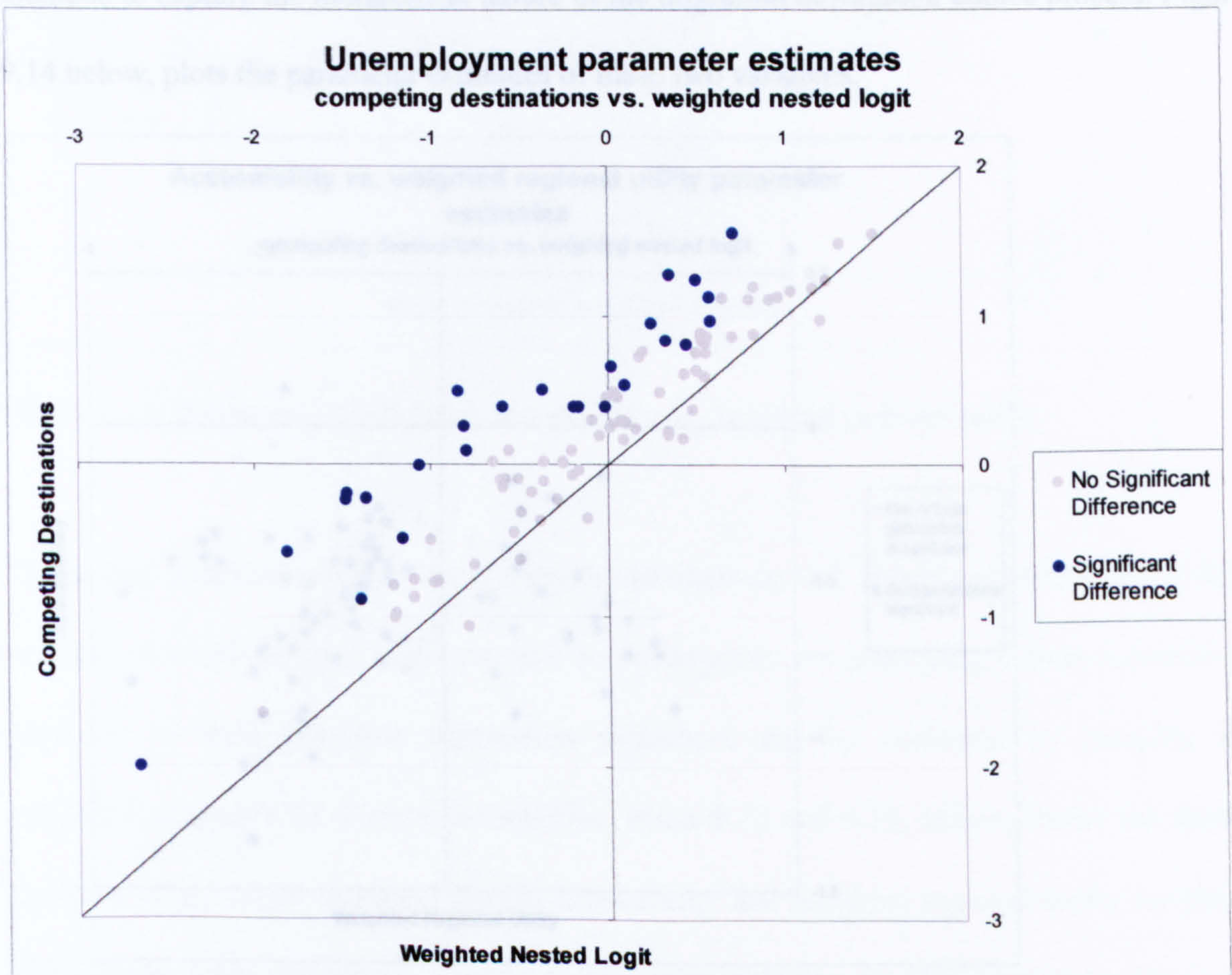


Figure 9.13: Unemp. parameter estimates: competing destinations vs. weighted NL

The most marked trends from figures 9.8 through 9.13 are the systematic difference between the two models' parameter estimates for the house price, tenure and unemployment variables. This is consistent with the finding, in chapter 7 above, that the competing destinations model's parameter estimates for these variables are affected by the multicollinearity in the model resulting from the correlations between the accessibility variable and the house price, tenure and unemployment variables. The similarities evident between figures 9.11, 9.12 and 9.13 and figures 7.12, 7.13 and 7.14 suggest that it is the same multicollinearity issues that affected the competing destinations parameter estimates that are influencing the hybrid models' parameter estimates for these variables.

It is also interesting to compare the parameter estimates for the two models' hierarchical variables. Whilst the competing destinations model's accessibility variable and the weighted nested logit model's regional utility variable are derived in very different ways, they are both intended to capture the hierarchical nature of the migration destination choice process. Figure 9.14 below, plots the parameter estimates of these two variables.

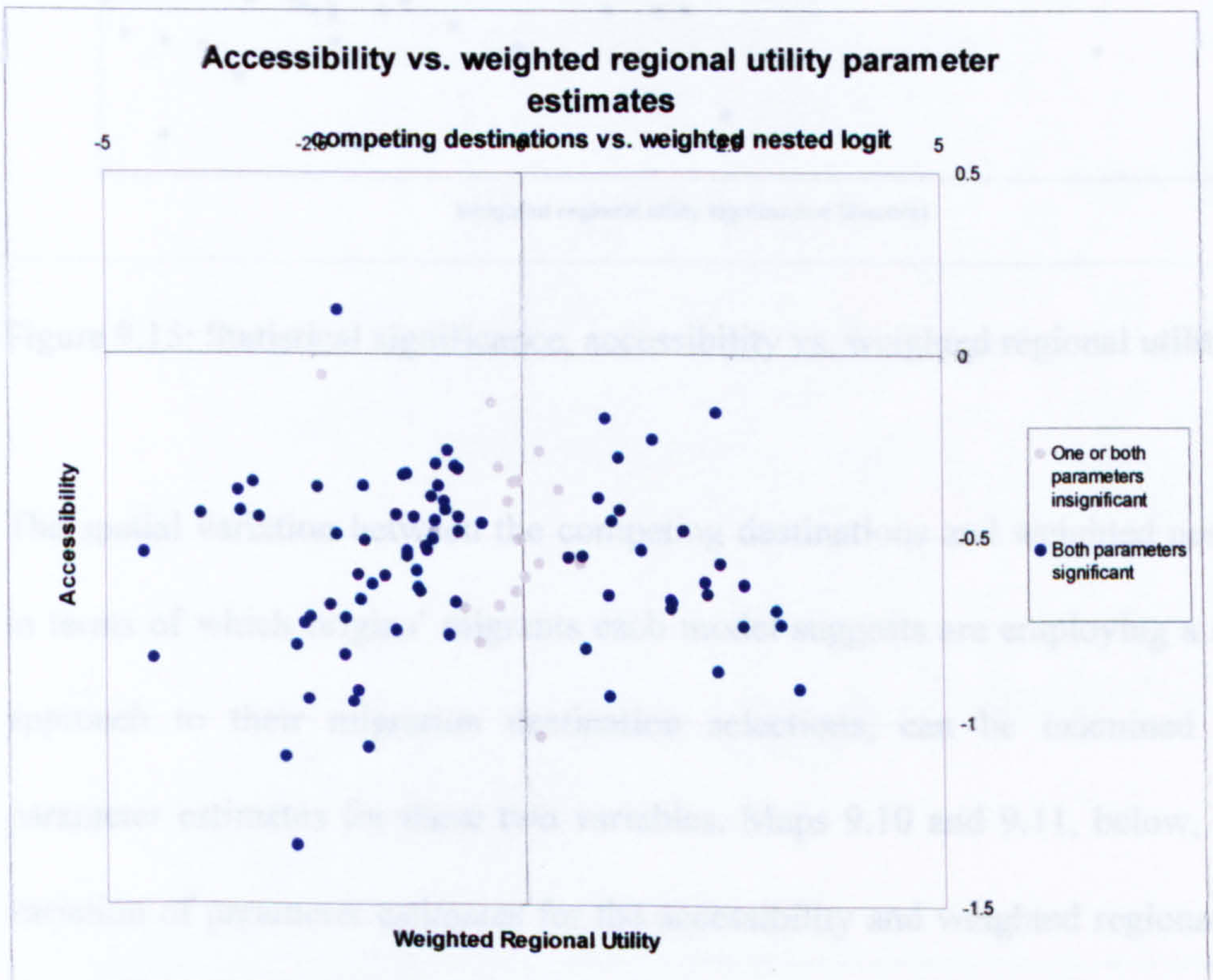


Figure 9.14: Accessibility vs. weighted regional utility parameter estimates

Of course there should be no expectation for values of accessibility and weighted regional utility to have similar values, given they are completely separate variables, with differing ranges of values and independent derivation. Figure 9.14, above, and a very low correlation coefficient of 0.026, confirm that there is indeed no such relationship between the values of these variables' parameter estimates. Furthermore, figure 9.15, below, confirms there is no correlation between those origins for which the two variables are more statistically significant.

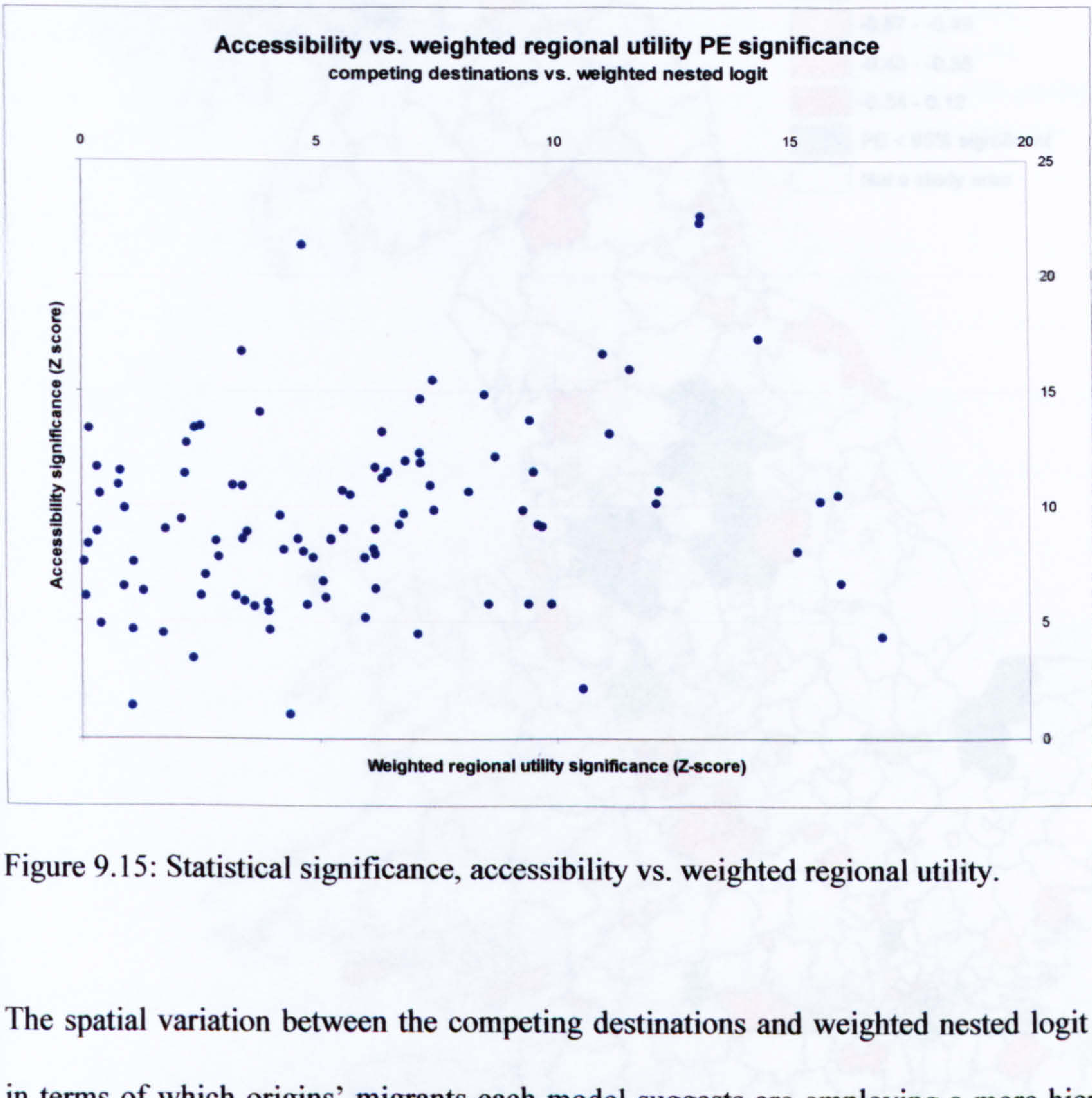
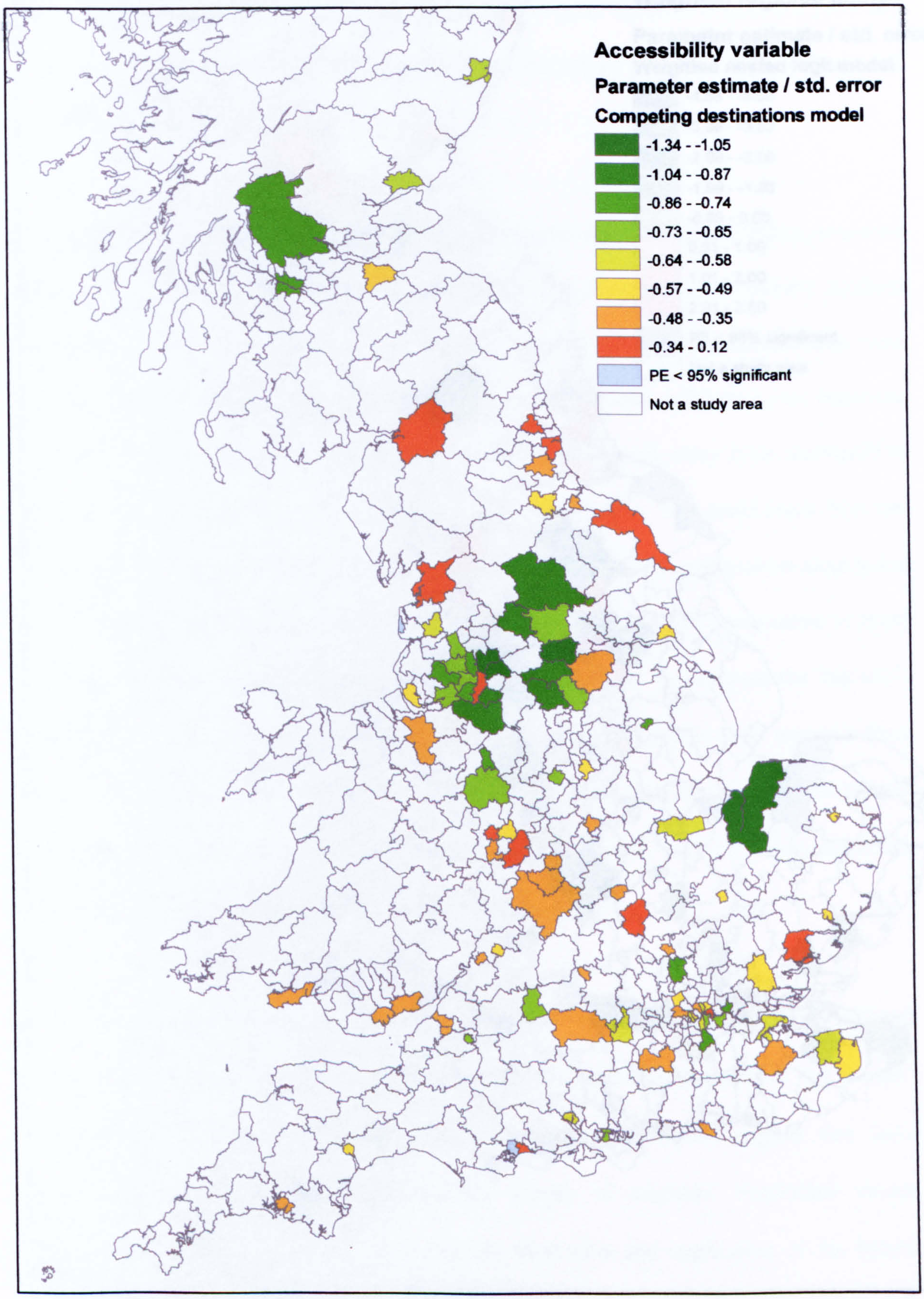


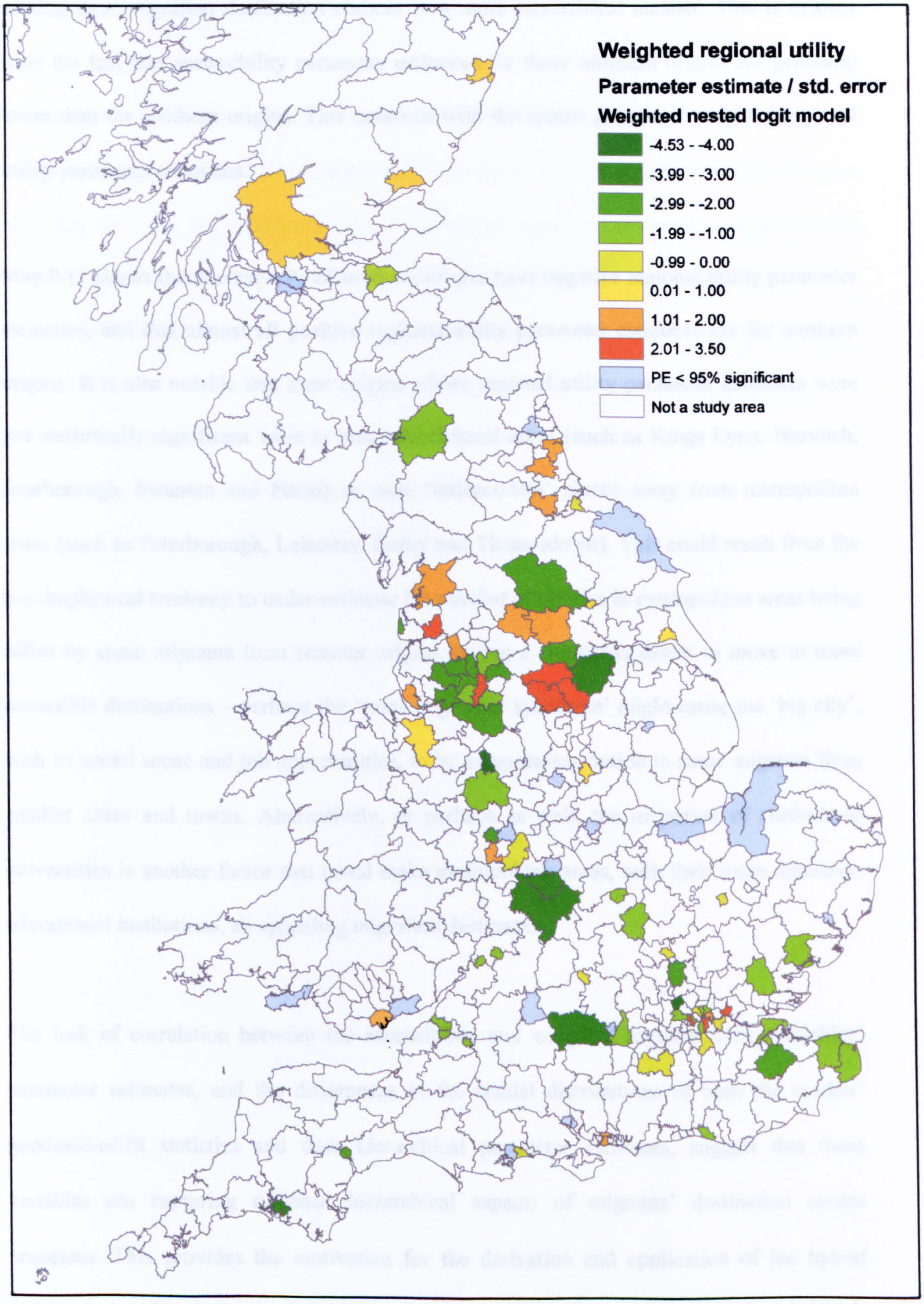
Figure 9.15: Statistical significance, accessibility vs. weighted regional utility.

The spatial variation between the competing destinations and weighted nested logit models, in terms of which origins' migrants each model suggests are employing a more hierarchical approach to their migration destination selections, can be examined by mapping the parameter estimates for these two variables. Maps 9.10 and 9.11, below, shows the spatial variation of parameter estimates for the accessibility and weighted regional utility variables, respectively. Only statistically significant parameter estimates (at 95% confidence level) are colour-coded on these maps - any study areas with parameter estimates that are less than 95%

significant are shaded light blue so that unreliable parameter estimate values do not misrepresent the actual distributions of hierarchical destination choice processing.



Map 9.10: Competing destinations accessibility parameter estimates, 95% sig. only.



Map 9.11: Weighted nested logit regional utility parameter estimates, 95% sig. only

It appears from map 9.10 that migrants from origins in the northern part of the country are making their migration destination choices in a more hierarchical manner. This is evident from the fact that accessibility parameter estimates for these northern origins are generally lower than for southern origins. This contrasts with the spatial pattern of weighted regional utility parameter estimates.

Map 9.11 shows that the majority of southern origins have negative regional utility parameter estimates, and that almost all positive regional utility parameter estimates are for northern origins. It is also notable that most origins whose regional utility parameter estimates were not statistically significant were in remote/peripheral areas (such as Kings Lynn, Norwich, Scarborough, Swansea and Poole) or were 'independent' places away from metropolitan areas (such as Peterborough, Leicester, Derby and Thamesdown). This could result from the psychophysical tendency to under-estimate the number of options in metropolitan areas being offset by some migrants from remoter origins having a conscious desire to move to more accessible destinations – perhaps the 'grass is greener syndrome' might cause the 'big city', with its social scene and job opportunities, to be an appealing option to some migrants from smaller cities and towns. Alternatively, or perhaps as well, the migration of students to universities is another factor that could make metropolitan areas, with their more numerous educational institutions, an appealing migration destination.

The lack of correlation between the accessibility and weighted regional utility variables' parameter estimates, and the differences in the spatial distributions of both the models' goodness-of-fit statistics and their hierarchical parameter estimates, suggest that these variables are capturing different hierarchical aspects of migrants' destination choice processes. This provides the motivation for the derivation and application of the hybrid weighted nested logit that includes both of these variables together in the same model. This hybrid model was introduced in chapter 5 and its results, presented above in chapter 8, can be seen to clearly justify its application.

Summary

This chapter has compared the goodness-of-fit statistics, error flow residuals and parameter estimates resulting from calibrations of a variety of migration destination choice models derived from hierarchical principles. Several systematic differences between the models have been highlighted and where possible have been explained in terms of the theoretical underpinnings of the various models. The various comparisons in this chapter have demonstrated that of all the hierarchical models applied here it is the hybrid weighted nested logit model that provides the most accurate destination choice predictions.

The next chapter applies this new hybrid weighted nested logit model to compare the migration destination choice behaviour of migrants grouped by age, gender and marital status.

Chapter Ten

Age, Gender and Marital Status Variation in Migration Destination Choice Behaviour

The three preceding chapters have examined variation in the goodness-of-fit and predicted parameter estimates from a variety of migration destination choice models, all of which have been calibrated against observed flows of all migrants aged 16 years and over. This was done in order to maximize sample size and thus minimize the standard errors of all parameter estimates.

This chapter compares the results of model calibrations against observed migration behaviour of migrant subgroups in order to highlight any variation in migration destination choice behaviour based on migrant age, gender or marital status. The preceding chapters have also established that the hybrid weighted nested logit model is the best performing model, in terms of goodness-of-fit to observed migration data. For this reason the examination of migrant subgroup behaviour in this chapter is based on calibrations of the hybrid weighted nested logit model.

Age Group Comparisons

In order to maintain sample size at a reasonable level, migrants were disaggregated only by broad age group: 16-24 (young adult), 25-54 (mid-life) and 55+ (approximately retirement age). Comparisons of the R^2_{adj} goodness-of-fit statistics from model calibrations for these three migrant groups are presented in figures 10.1 to 10.6 below. The AIC goodness-of-fit statistic is not used for comparative purposes between migrant subgroups because its value is partially based upon the inherent variance in the observed data under consideration and is therefore not comparable between models calibrated against different observed data. It should be mentioned that a fundamental assumption of all the migration models calibrated in this research is the independence of each individual migrant. However, this assumption is clearly not valid for all migrants, as many couples and families migrate as a group (Boyle *et al*, 2001). Up to this point any inaccuracy due to relaxation of this assumption has affected all model calibrations equally. However, it should be noted that some of the results presented in this chapter are inherently more susceptible to the affects of joint-migrations. Specifically, the married and, to a lesser degree, the 25-54 year old migrant subgroups are likely to exhibit more joint migrations than other subgroups. It is not simple to assess the impact of this problem, and the data does not permit joint migrations to be isolated and modelled independently, so it is the authors hope that any inaccuracies due to joint migrations will be negligible, especially given the larger standard errors of parameter estimates presented in this chapter due to the migrant sub-groups' smaller sample sizes.

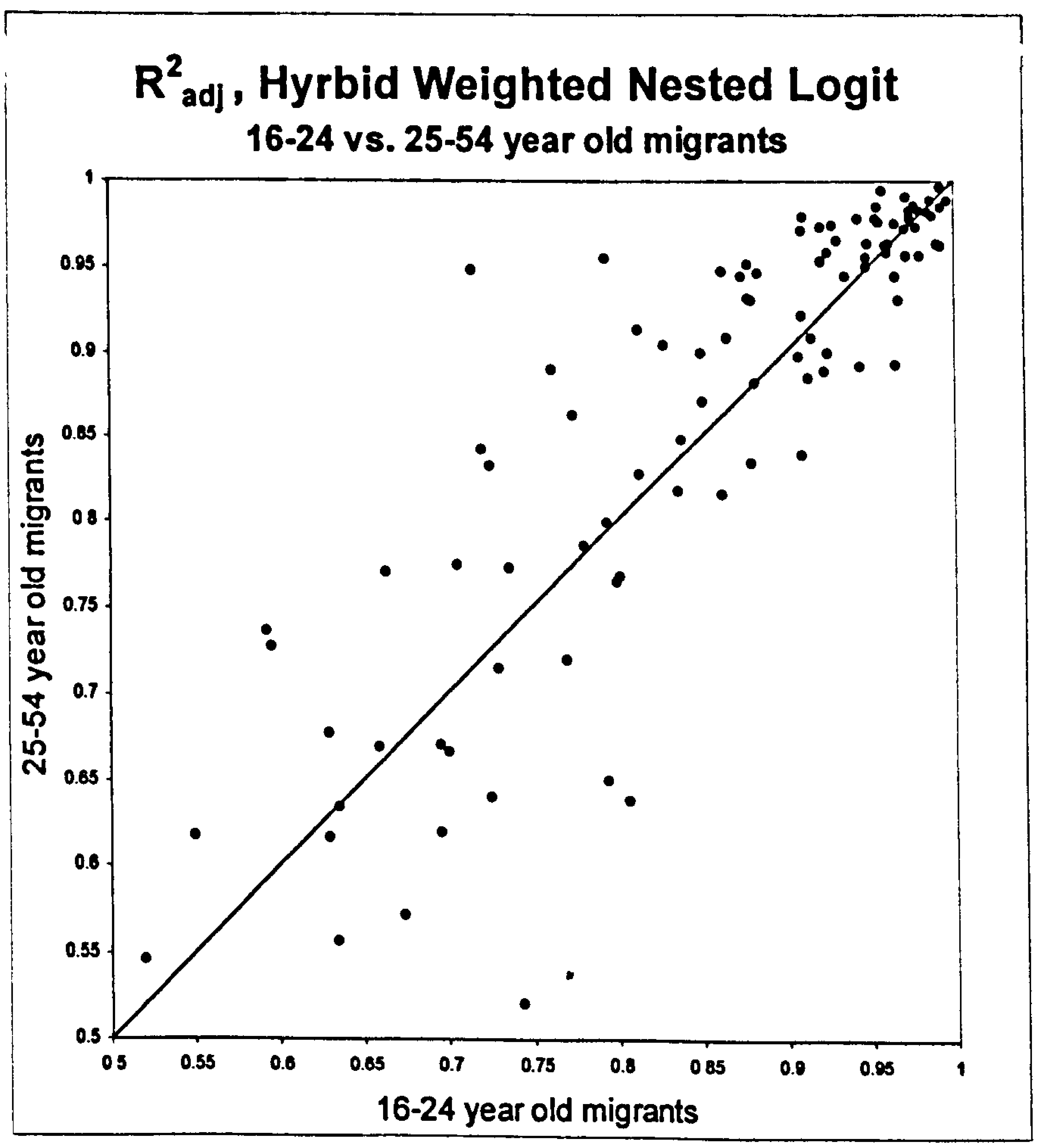


Figure 10.1: R^2_{adj}, migrants aged 16-24 vs. 25-54, hybrid weighted nested logit.

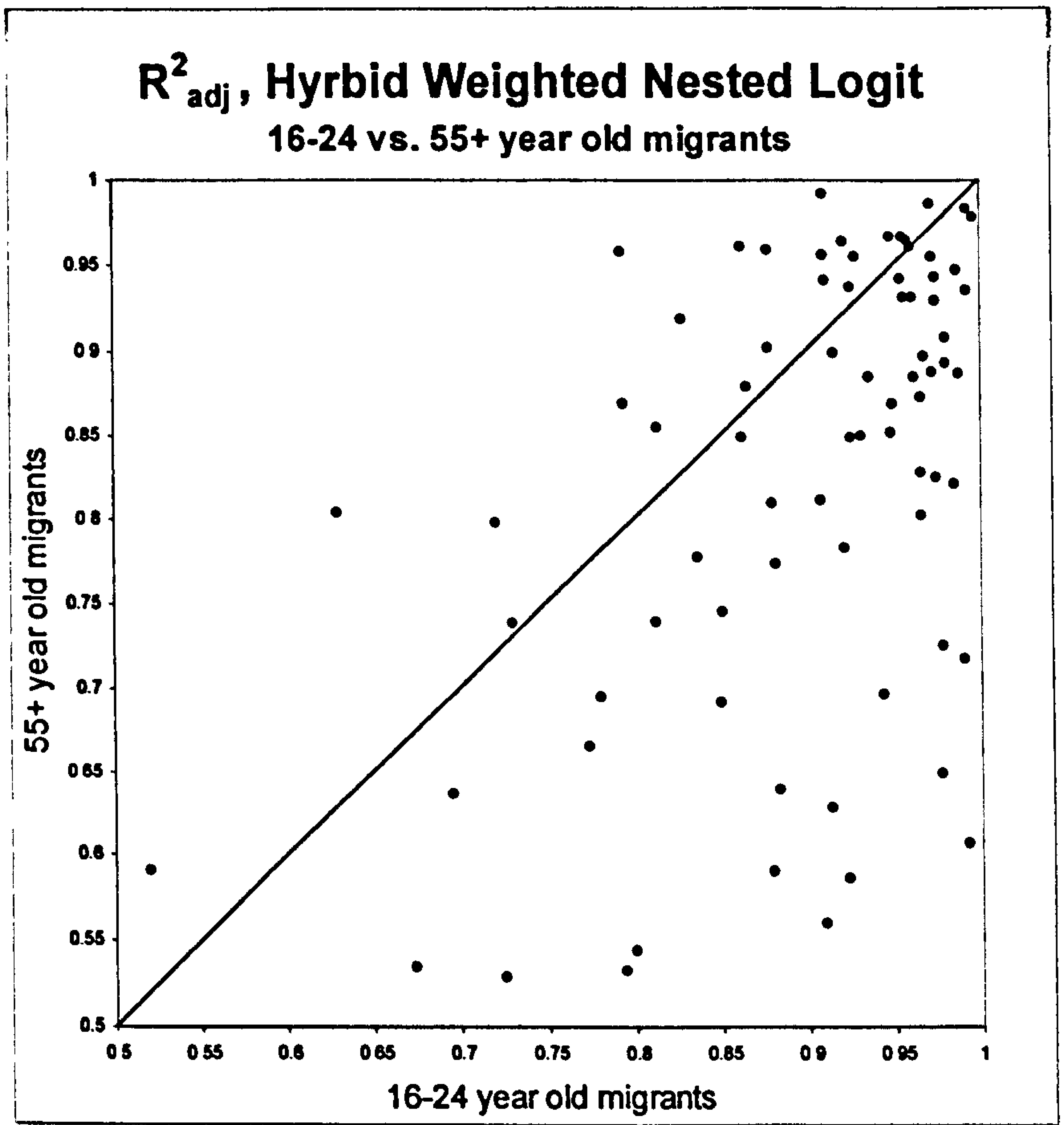


Figure 10.2: R^2_{adj}, migrants aged 16-24 vs. 55+, hybrid weighted nested logit.

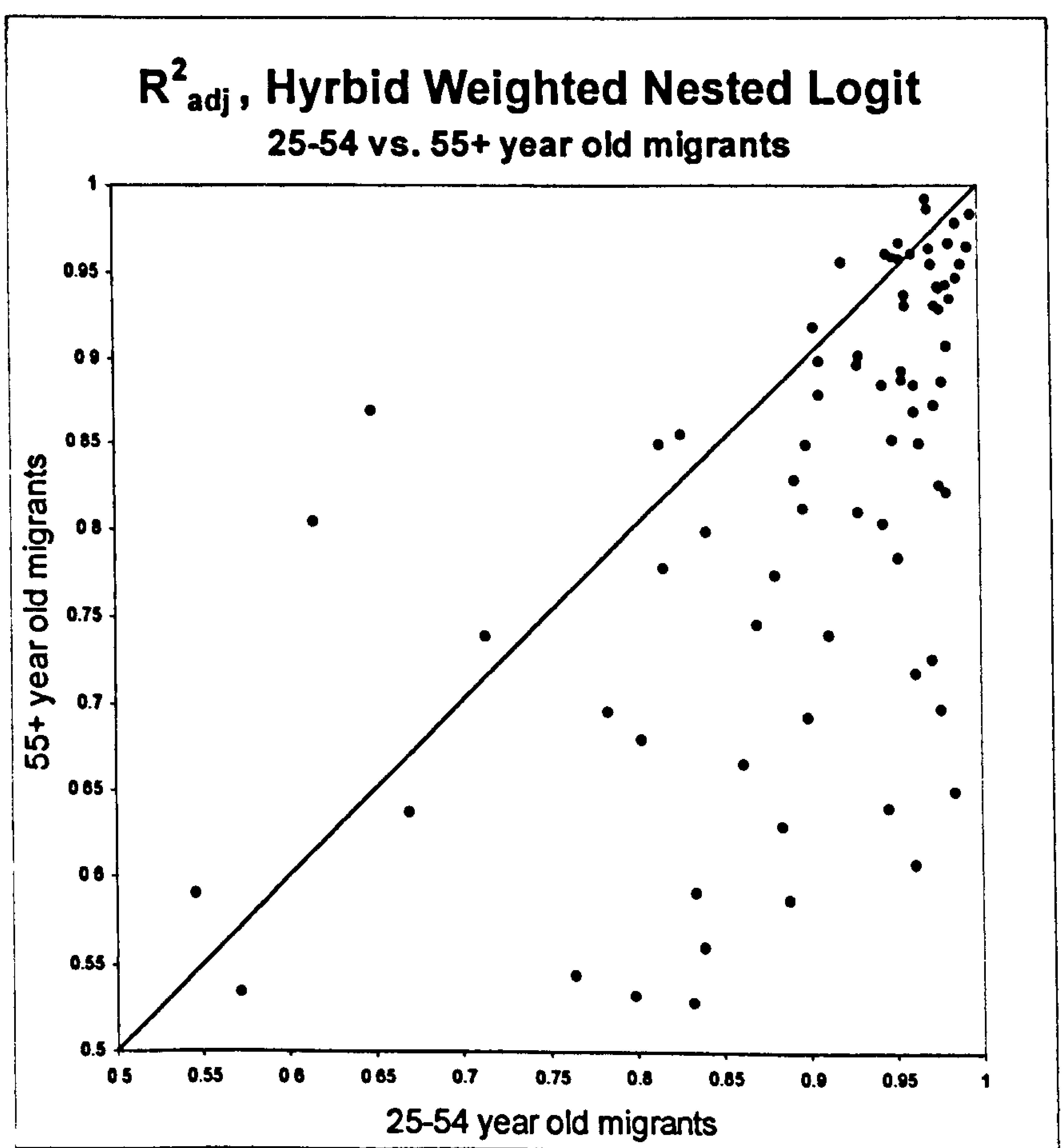


Figure 10.3: R^2_{adj}, migrants aged 25-54 vs. 55+, hybrid weighted nested logit.

When one considers the sample sizes of the migrant subgroups compared in figure 10.1 to 10.3, it becomes apparent that the majority of the variation in goodness-of-fit between the migrant age groups is the result of different numbers of migrants in each age group leaving the various migration origins. The average out-flows of each of these three broad migrant age-groups from the 100 sample origins are presented in table 10.1, below.

Migrant age group	Average gross out-flow
16-24	744.41
25-54	1077.88
55+	147.69

Table 10.1: Average gross out-migration per origin, by broad migrant age group.

The average out-flows presented in table 10.1 correlate with the R^2_{adj} goodness-of-fit figures seen above. The 25-54 year-old group has the best goodness-of-fit followed by 16-24 and then 55+ year old migrants. Given this influence on the R^2_{adj} numbers, no meaningful conclusions can be drawn from figures 10.1 through 10.3. However, it is possible to compare the goodness-of-fit of different models calibrated for the same migrant subgroup. So, for instance, in order to investigate age-group variation in 'how hierarchically' migrants select their destinations, age group variation in extent of the hybrid model's goodness-of-fit improvement over the traditional model can be compared – see figures 10.4, 10.5 and 10.6, below.

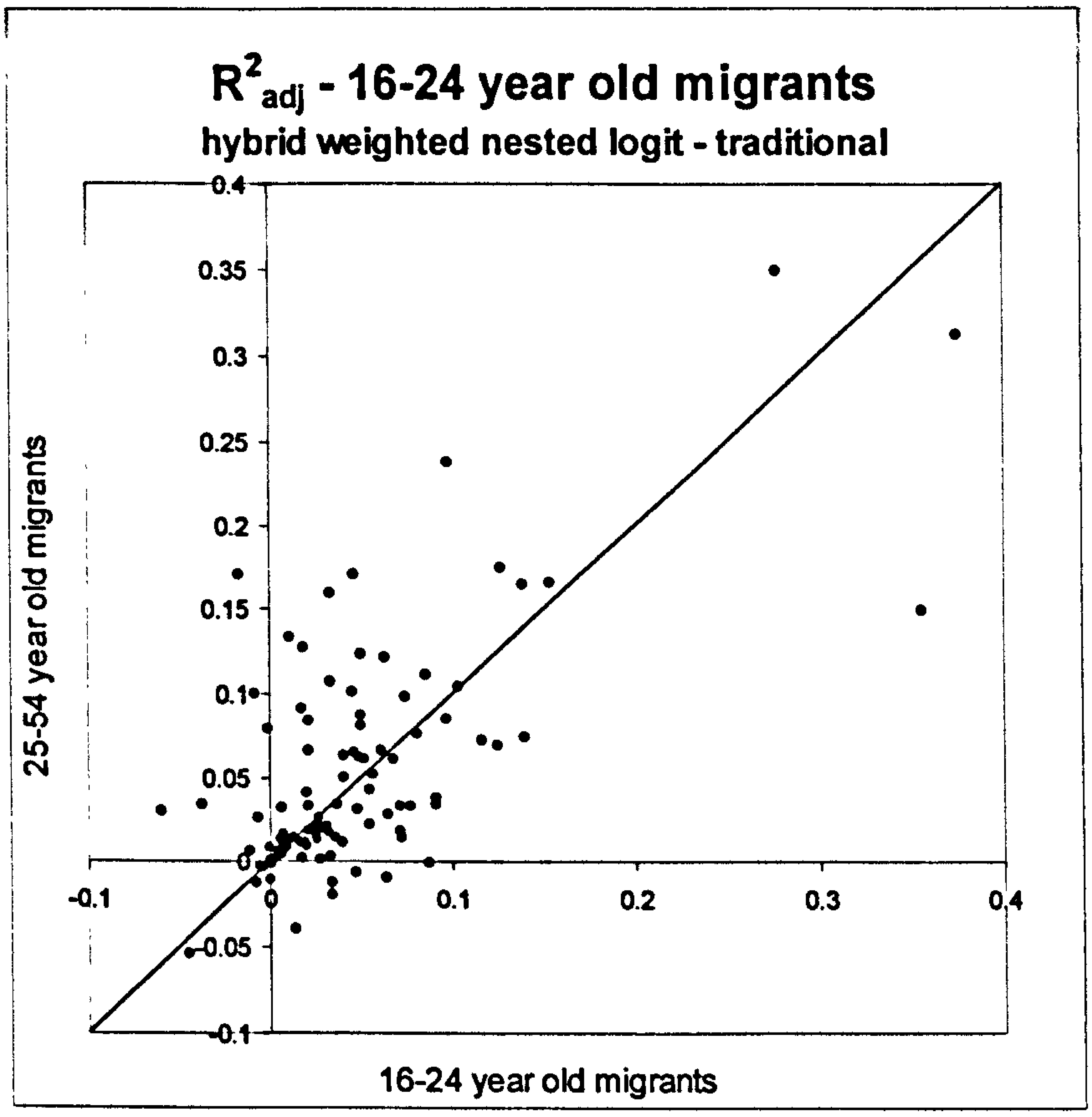


Figure 10.4: R^2_{adj} improvement, hybrid over traditional, 16-24 year old migrants.

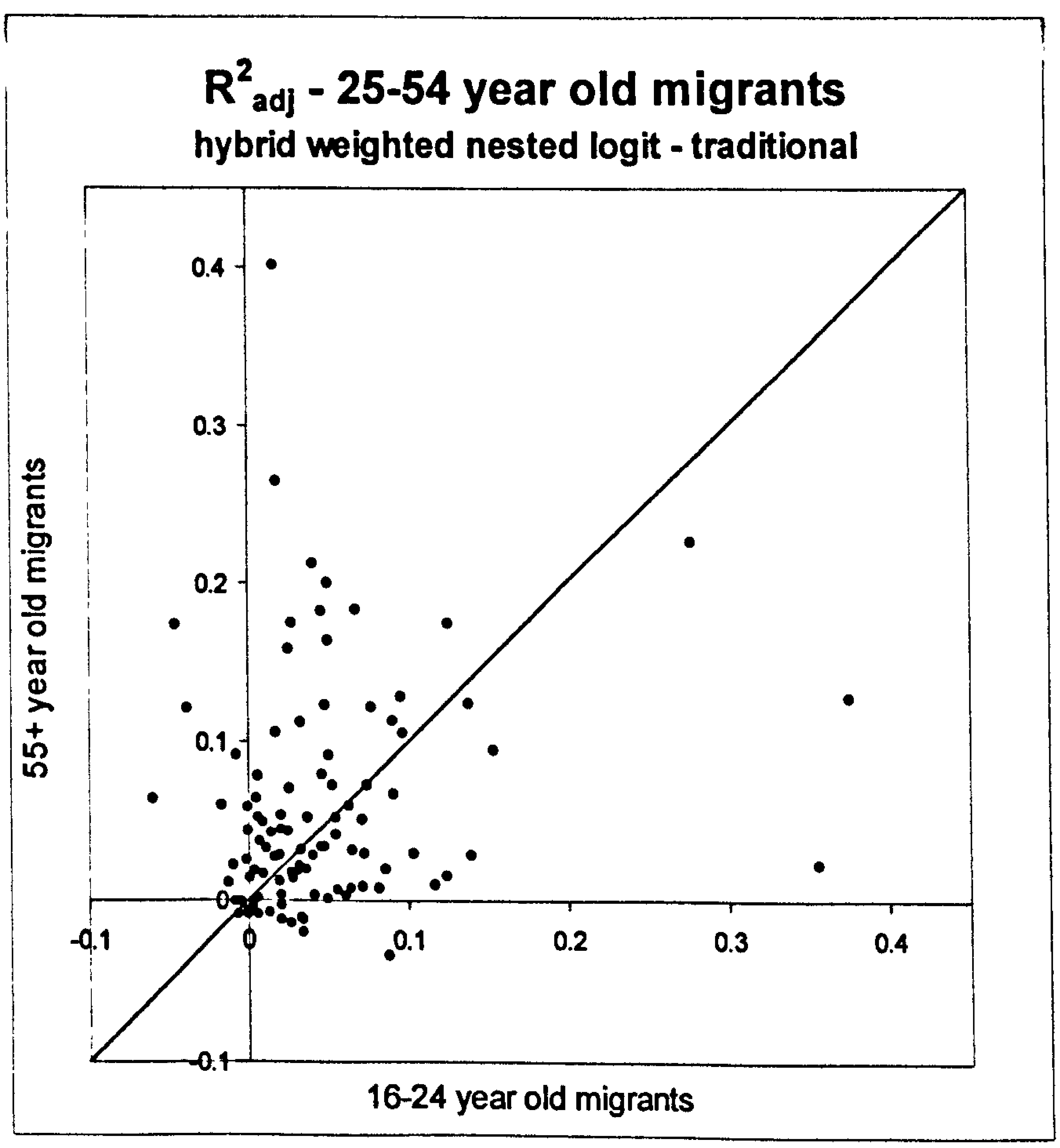


Figure 10.5: R^2_{adj} improvement, hybrid over traditional, 25-54 year old migrants.

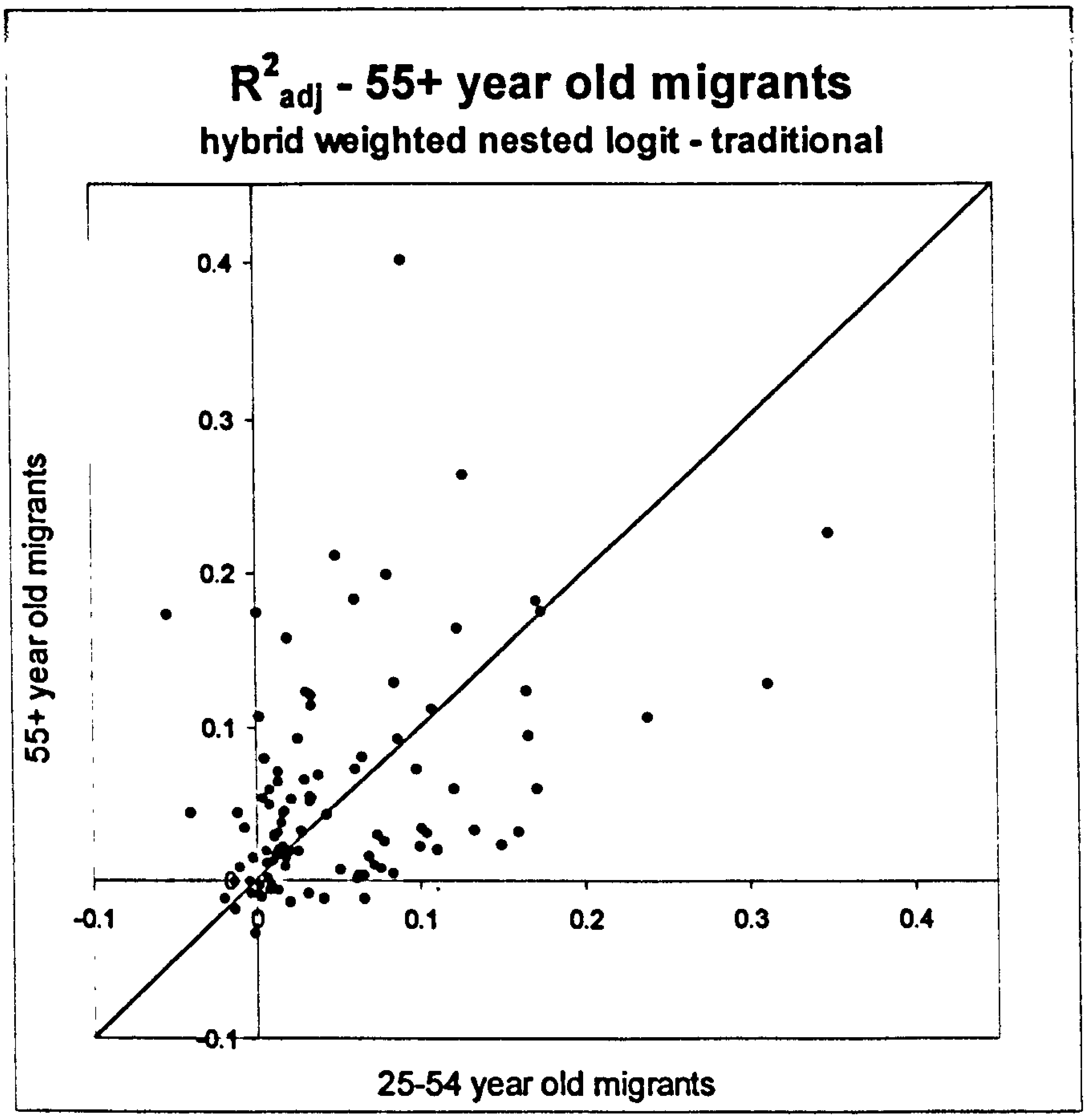


Figure 10.6: R^2_{adj} improvement, hybrid over traditional, 55+ year old migrants.

Figures 10.4 to 10.6 indicate that whilst there is a wide variety of goodness-of-fit improvement between the 100 selected migration origins, there is no systematic age-group variation, suggesting that in general both age groups benefit equally from the inclusion of the additional hierarchically-derived variables in the hybrid weighted nested logit model.

It is also interesting to compare some of the parameter estimates predicted for the different age groups. Most of the 24 comparisons possible between the eight explanatory variables of the hybrid model for the three combinations of age-groups, show no clear patterns – indeed many show very few origins with statistically significant differences between their parameter estimates. The more interesting comparisons are presented in figures 10.7 through 10.13 below.

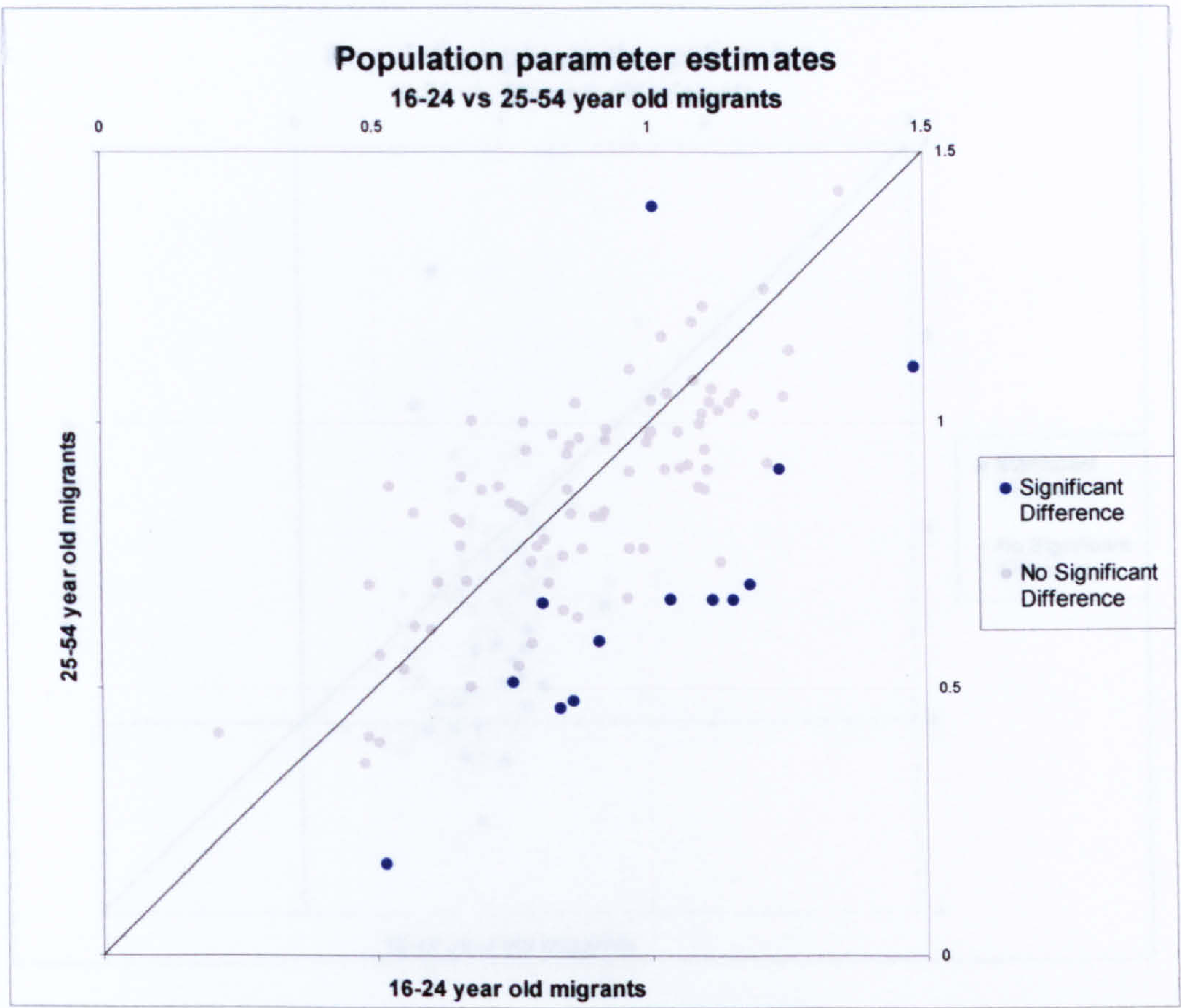


Figure 10.7: Population parameter estimates, 16-24 vs. 25-54 year old migrants.

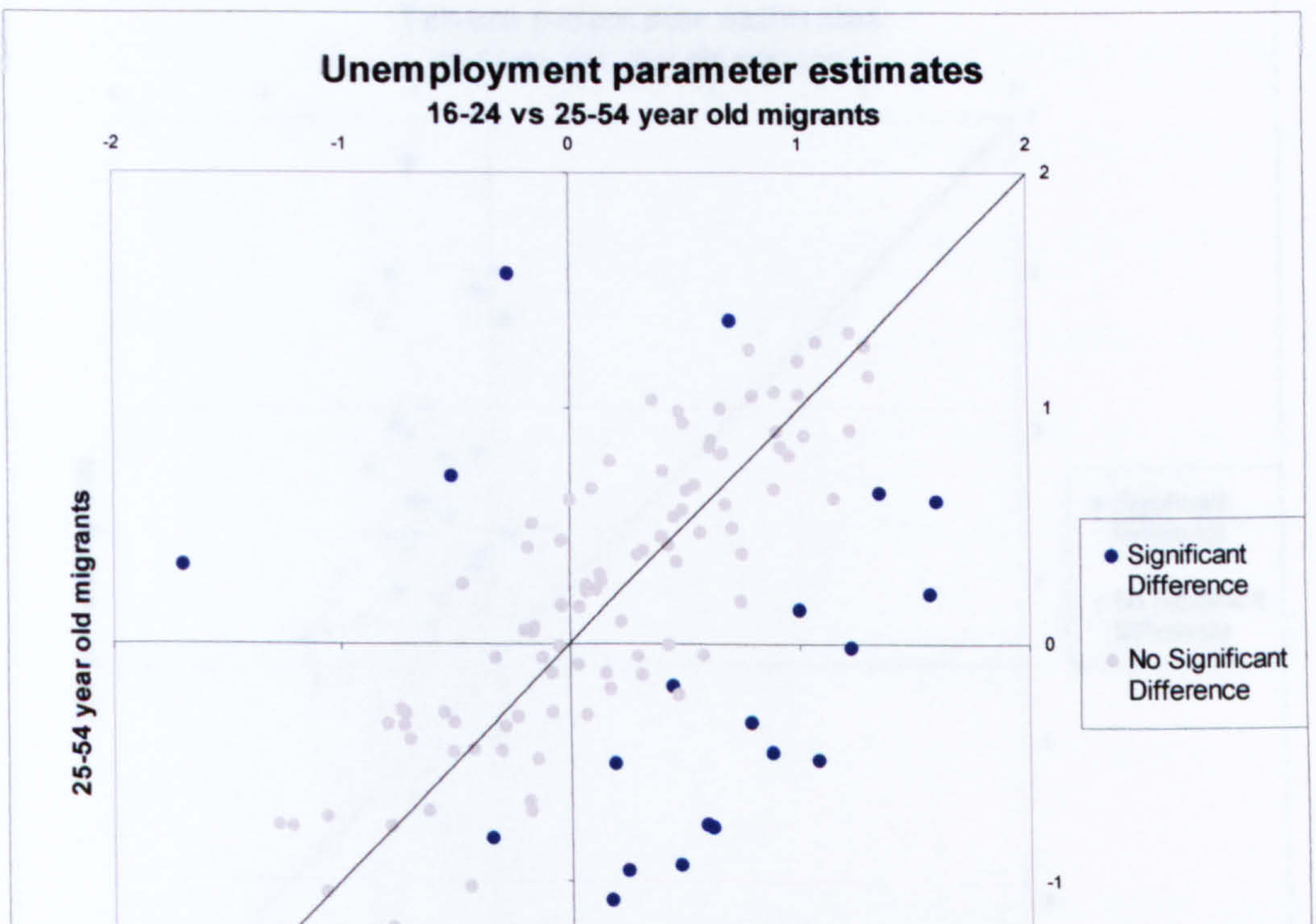


Figure 10.8: Unemployment parameter estimates, 16-24 vs. 25-54 year old migrants.

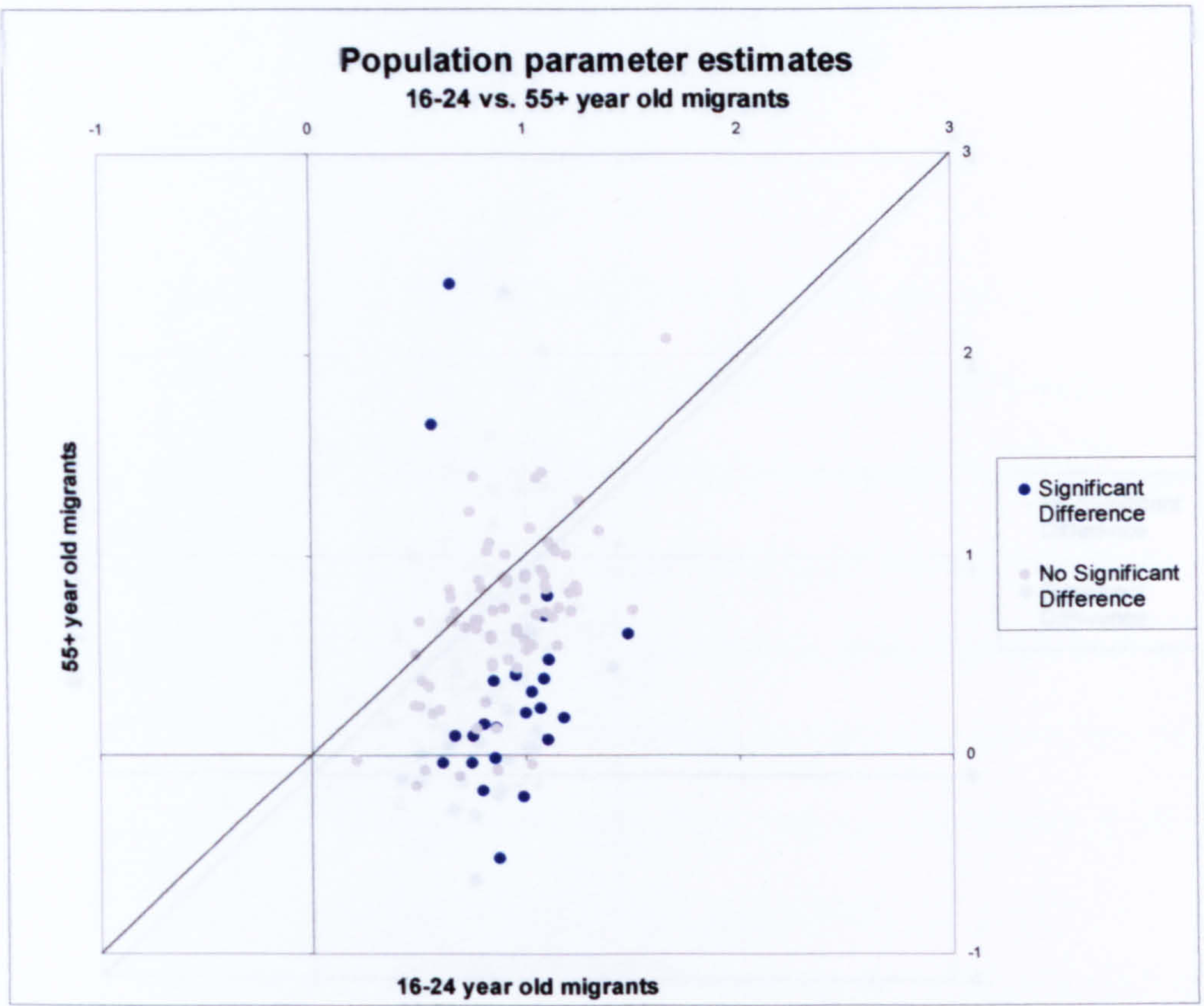


Figure 10.9: Population parameter estimates, 16-24 vs. 55+ year old migrants.

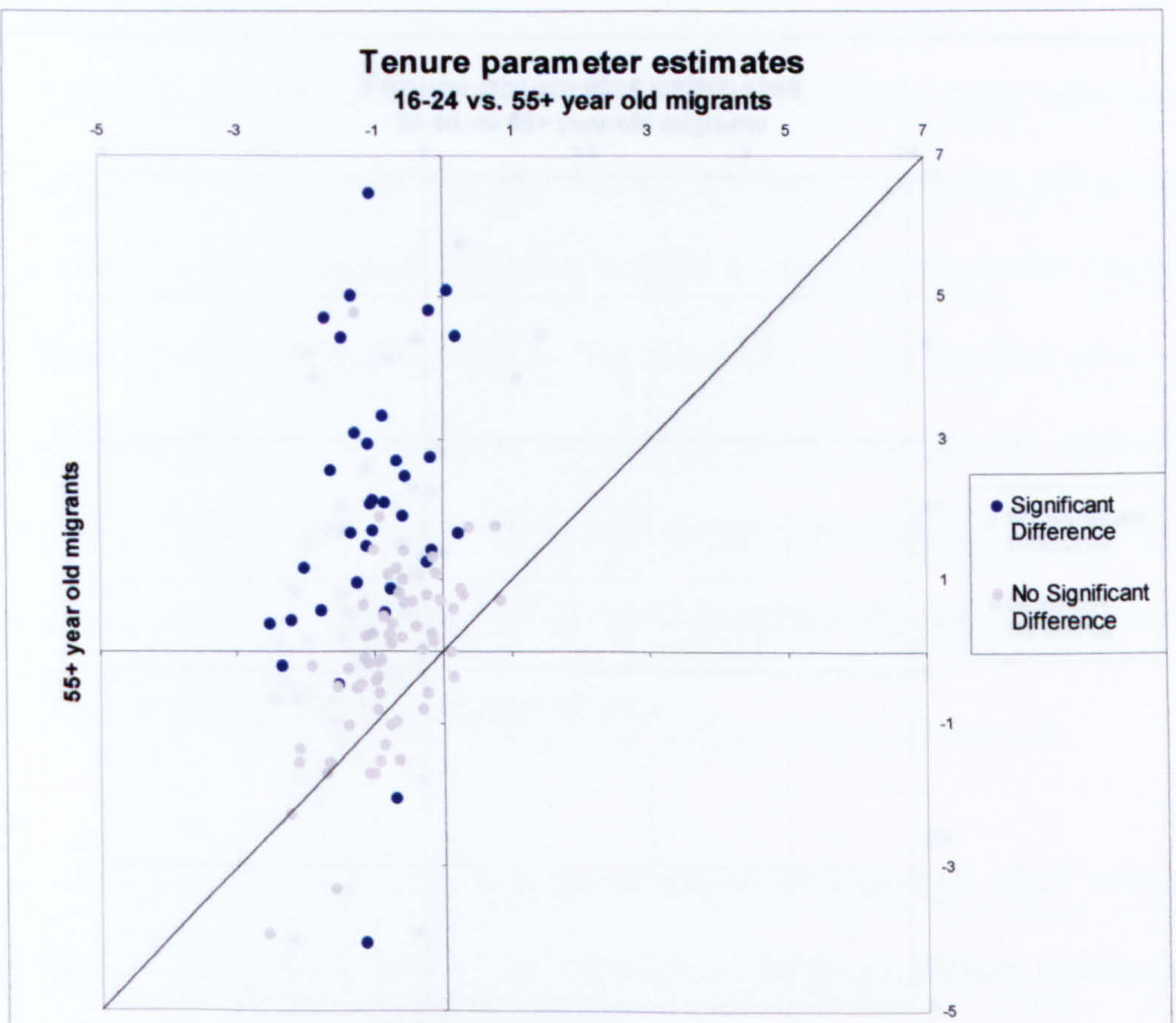


Figure 10.10: Tenure parameter estimates, 16-24 vs. 55+ year old migrants.

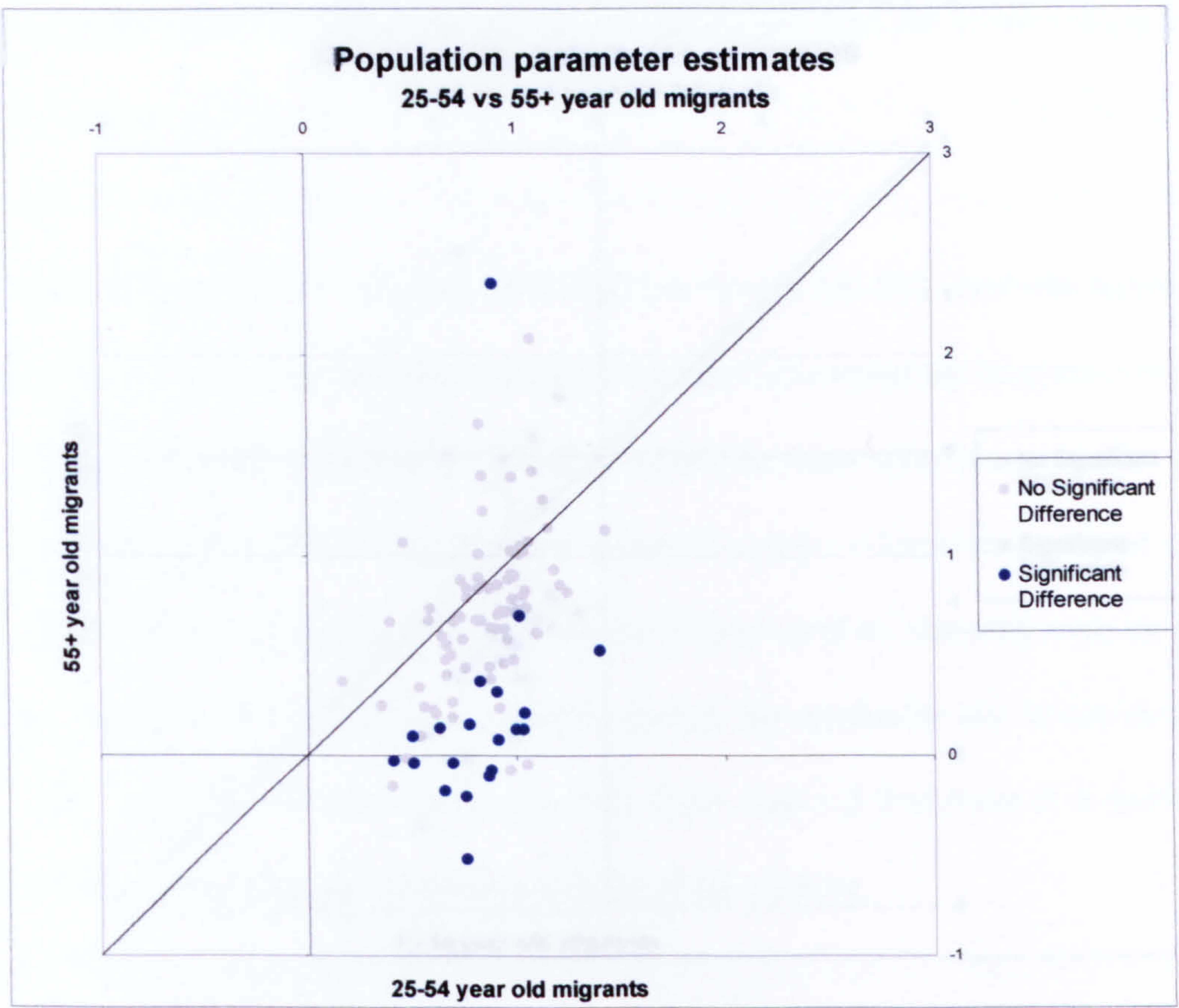


Figure 10.11: Population parameter estimates, 25-54 vs. 55+ year old migrants.

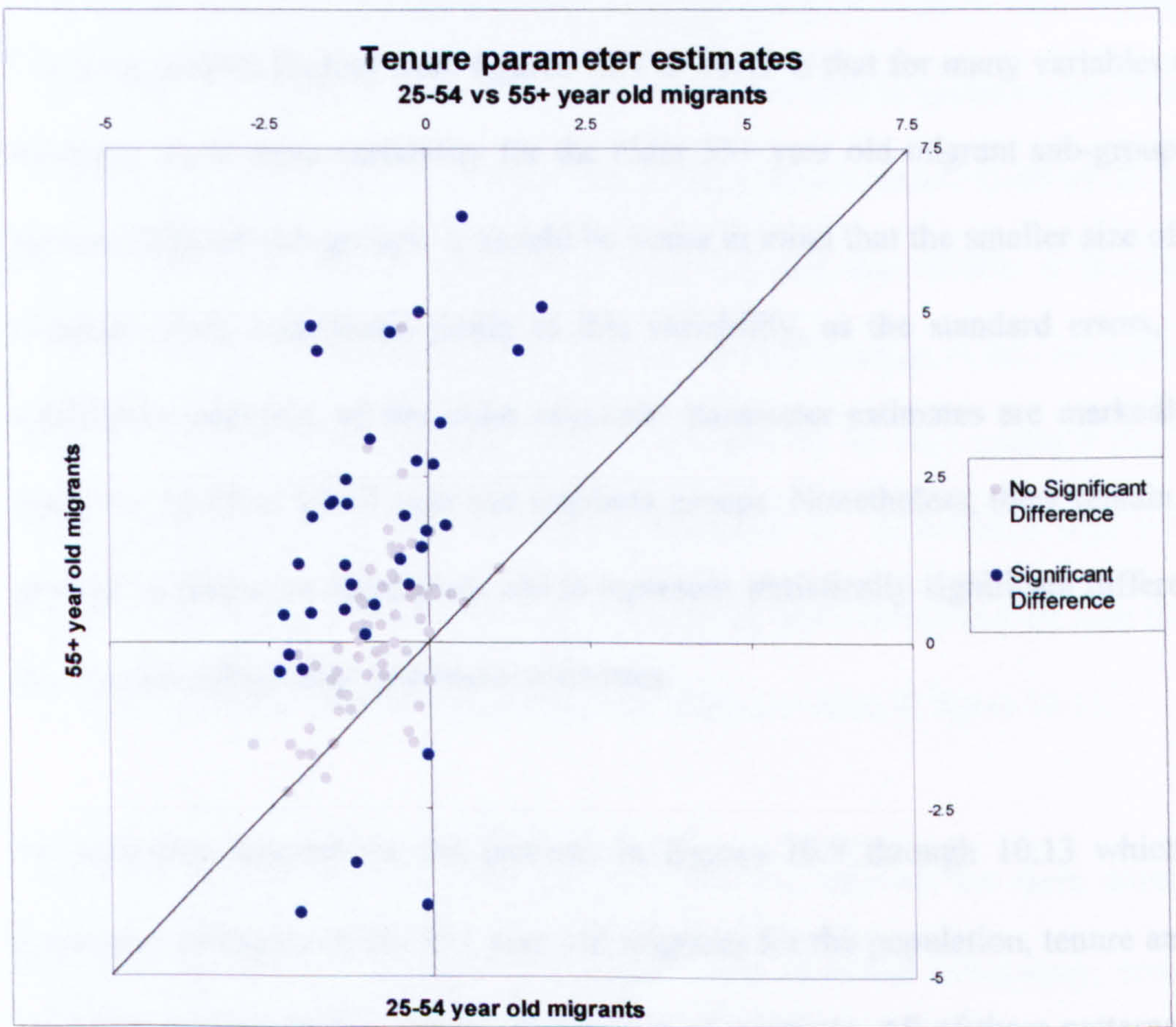


Figure 10.12: Tenure parameter estimates, 25-54 vs. 55+ year old migrants.

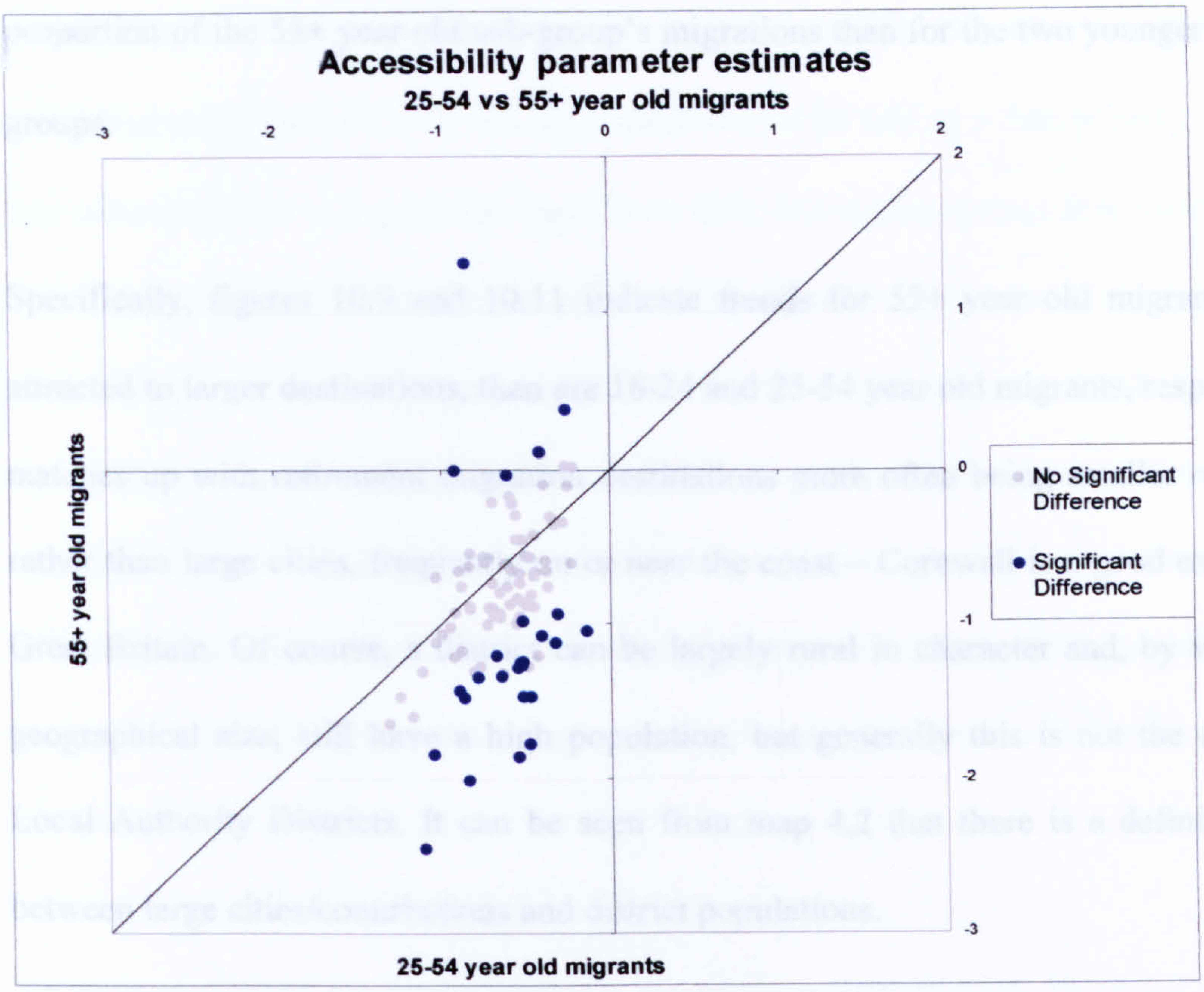


Figure 10.13: Accessibility parameter estimates, 25-54 vs. 55+ year old migrants.

The most notable finding from figures 10.7 to 10.13 is that for many variables the parameter estimates show more variability for the older 55+ year old migrant sub-group, than for the younger migrant sub-groups. It should be borne in mind that the smaller size of this group of migrants likely contributes partly to this variability, as the standard errors, and therefore confidence intervals, of the older migrants' parameter estimates are markedly higher than those for 16-24 or 25-54 year old migrants groups. Nonetheless, there remain a worthwhile number of points on these plots which represent statistically significant differences between the migrant sub-groups' parameter estimates.

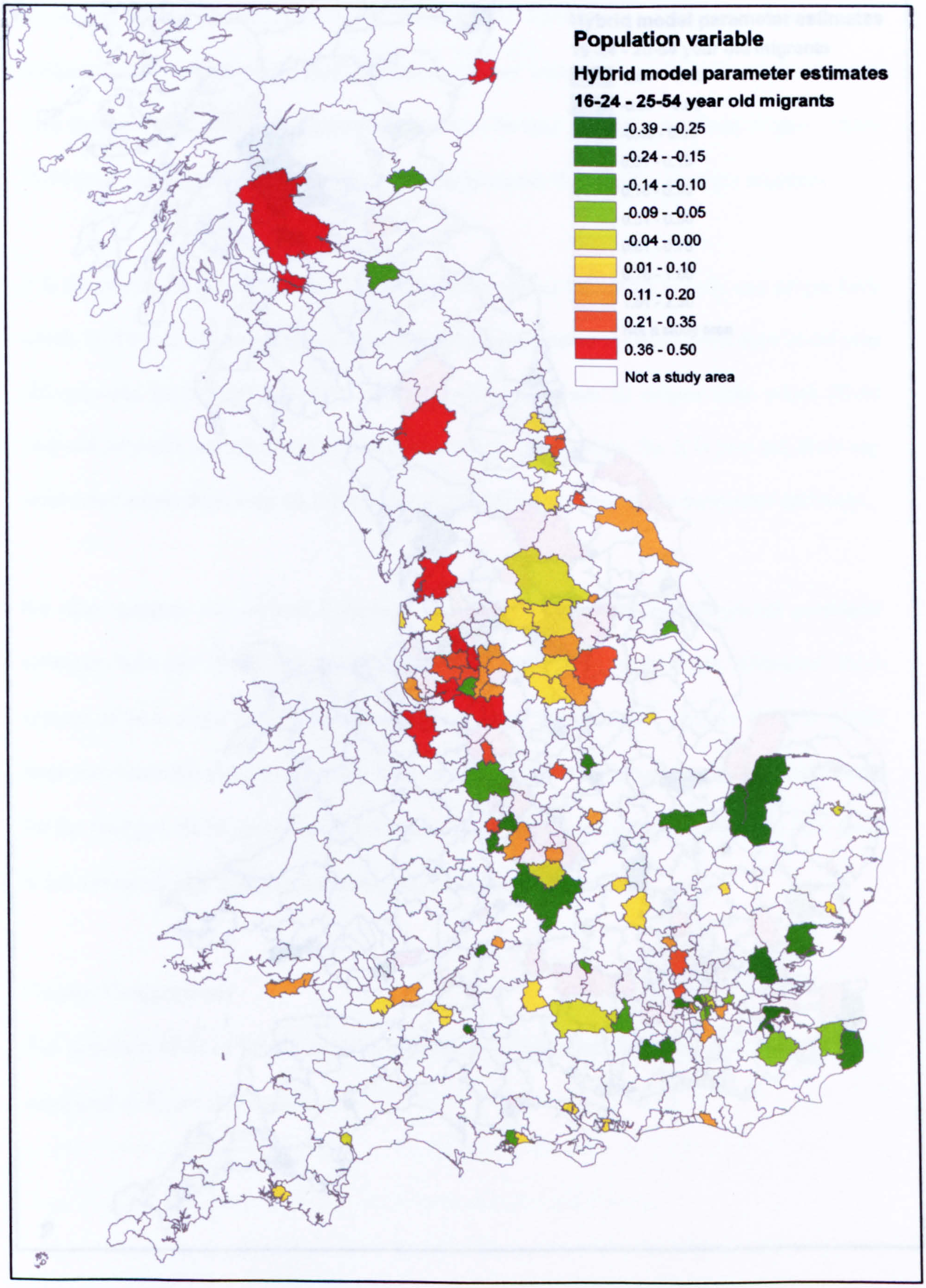
Of particular interest are the patterns in figures 10.9 through 10.13 which compare the parameter estimates of the 55+ year old migrants for the population, tenure and accessibility variables with those for younger age groups of migrants. All of these patterns are consistent with typical retirement migration behaviour, which one would expect to account for a larger proportion of the 55+ year old sub-group's migrations than for the two younger migrant sub-groups.

Specifically, figures 10.9 and 10.11 indicate trends for 55+ year old migrants to be less attracted to larger destinations, than are 16-24 and 25-54 year old migrants, respectively. This matches up with retirement migration destinations more often being smaller rural locations rather than large cities, frequently on or near the coast – Cornwall is a good example within Great Britain. Of course, a district can be largely rural in character and, by virtue of large geographical size, still have a high population, but generally this is not the case for 1991 Local Authority Districts. It can be seen from map 4.2 that there is a definite correlation between large cities/conurbations and district populations.

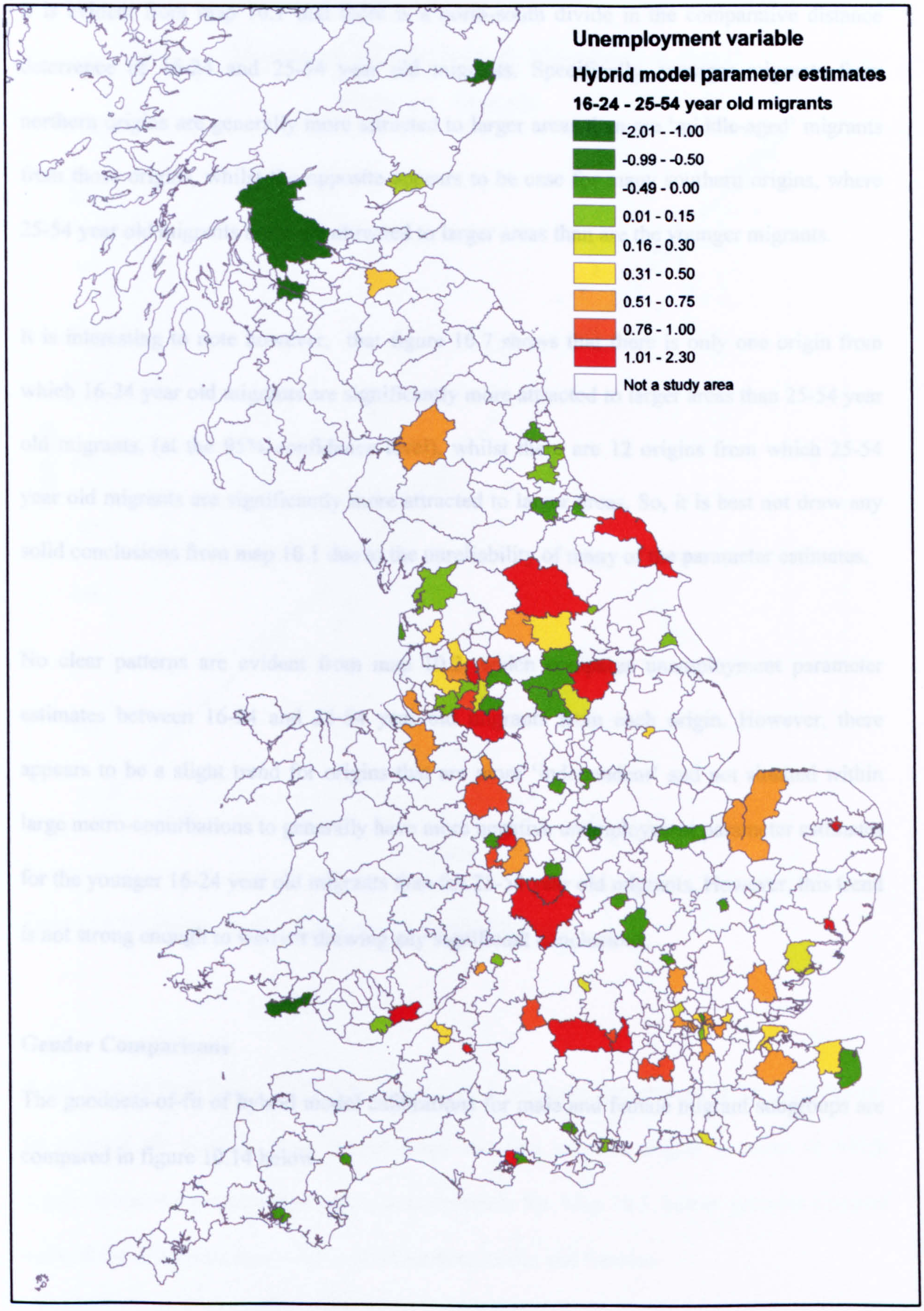
Figure 10.13 indicates that 55+ year old migrants from some origins are significantly more deterred from moving to areas with higher accessibility, than 25-54 year old migrants. It will be recalled from the derivation of the accessibility variable in chapter 6 that it is a measure of how close a destination is to how many other larger areas. Derived from hierarchical assumptions of the migration destination choice process, accessibility is intended to represent the likelihood of a destination being cognized within a larger cluster of destinations, rather than in isolation or in a smaller cluster. However, by its very nature it also differentiates effectively between how urban or rural a destination is – destinations with low accessibility are generally more isolated. Thus low accessibility also correlates well with the most typical retirement migration destinations, hence the pattern evident in figure 10.13.

Figures 10.10 and 10.12 indicate that 55+ year old migrants are more attracted to or less deterred from moving to areas with higher rates of owner occupancy, compared to both younger migrant sub-groups. It is likely that this results from older migrants generally being more well-established socio-economically than younger migrants, or in the colloquial parlance, to have been 'settled down' for some time. Older migrants are more likely to be more financially secure - with the costs associated with raising a family being a thing of the past older migrants will generally have lower debt and higher savings than younger migrants. This means that such older migrants are more likely to already be owner occupiers, and once on the 'property ladder' most people tend to stay there, though many decide to free up some equity by moving to smaller, more manageable properties at some point. Thus, the majority of older migrants will be in search of another property to purchase when they migrate, and will be attracted to areas with higher rates of owner-occupancy which typically have more properties available for purchase than will be the case for areas with lower rates of owner occupancy.

As discussed above, the patterns observed in figures 10.9 through 10.13 are believed to be caused by the particular migration destination choices of retirement migrants. The cause of the patterns in figures 10.7 and 10.8 are less clear, so maps 10.1 and 10.2, below, attempt to shed additional light on this matter by plotting the spatial variation in these age groups' parameter estimate differences. Specifically, map 10.1 presents the differences in population parameter estimates between 16-24 and 25-54 year old migrants, and map 10.2 presents differences in unemployment parameter estimates between 16-24 and 25-54 year old migrants.



Map 10.1: Difference in population parameter estimates, 16-24 – 25-54 year olds.



Map 10.2: Difference in unemployment parameter estimates, 16-24 – 25-54 year olds.

It is evident from map 10.1 that there is a north-south divide in the comparative distance deterrence of 16-24 and 25-54 year old migrants. Specifically, younger migrants from northern origins are generally more attracted to larger areas than are 'middle-aged' migrants from those origins, whilst the opposite appears to be case for many southern origins, where 25-54 year old migrants are more attracted to larger areas than are the younger migrants.

It is interesting to note however, that figure 10.7 shows that there is only one origin from which 16-24 year old migrants are significantly more attracted to larger areas than 25-54 year old migrants, (at the 95% confidence level), whilst there are 12 origins from which 25-54 year old migrants are significantly more attracted to larger areas. So, it is best not draw any solid conclusions from map 10.1 due to the unreliability of many of the parameter estimates.

No clear patterns are evident from map 10.2, which compares unemployment parameter estimates between 16-24 and 25-54 year old migrants from each origin. However, there appears to be a slight trend for origins that are more 'independent' and not situated within large metro-conurbations to generally have more negative unemployment parameter estimates for the younger 16-24 year old migrants than for 25-54 year old migrants. However, this trend is not strong enough to warrant drawing any significant conclusions.

Gender Comparisons

The goodness-of-fit of hybrid model calibrations for male and female migrant subgroups are compared in figure 10.14 below.

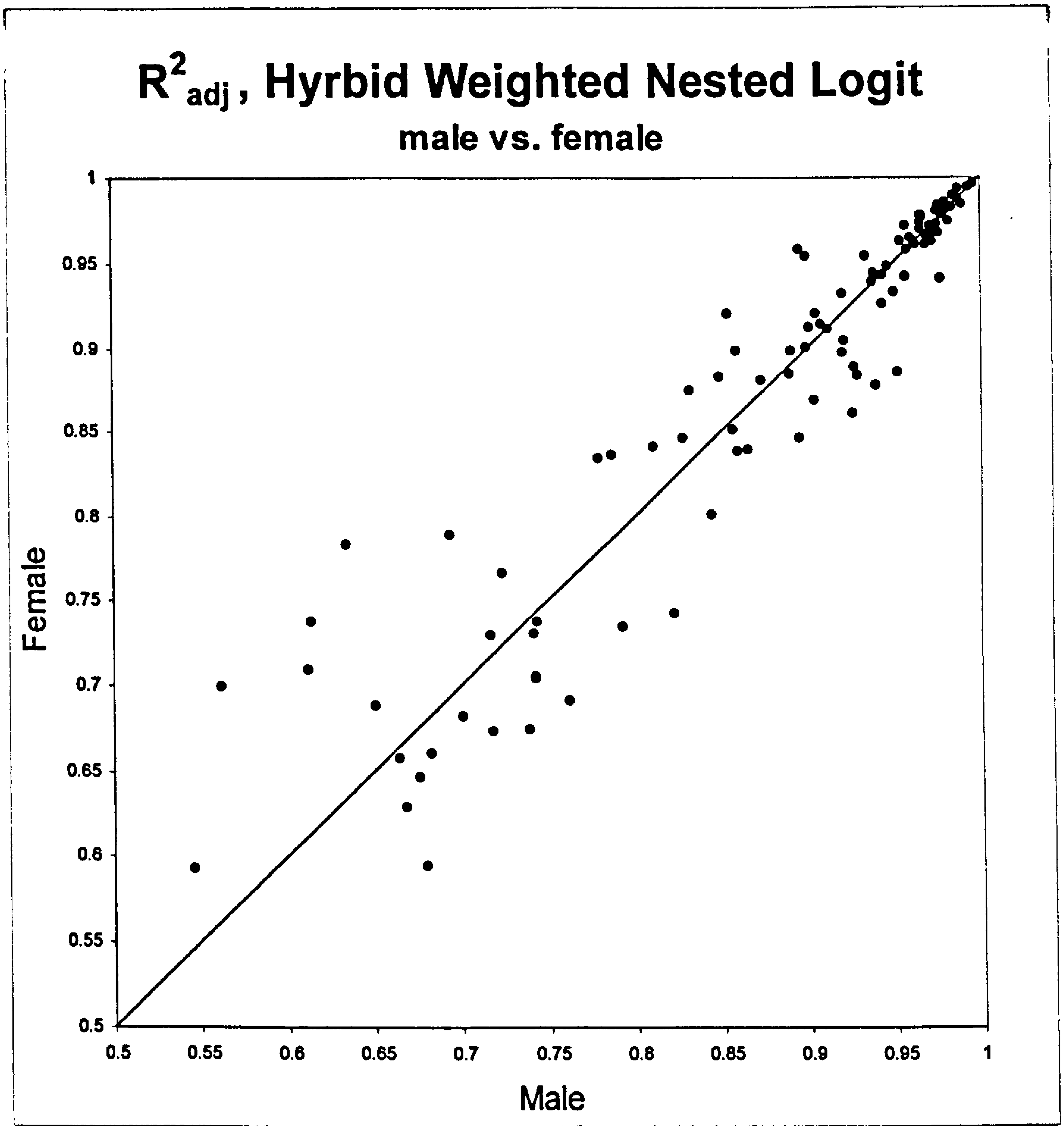
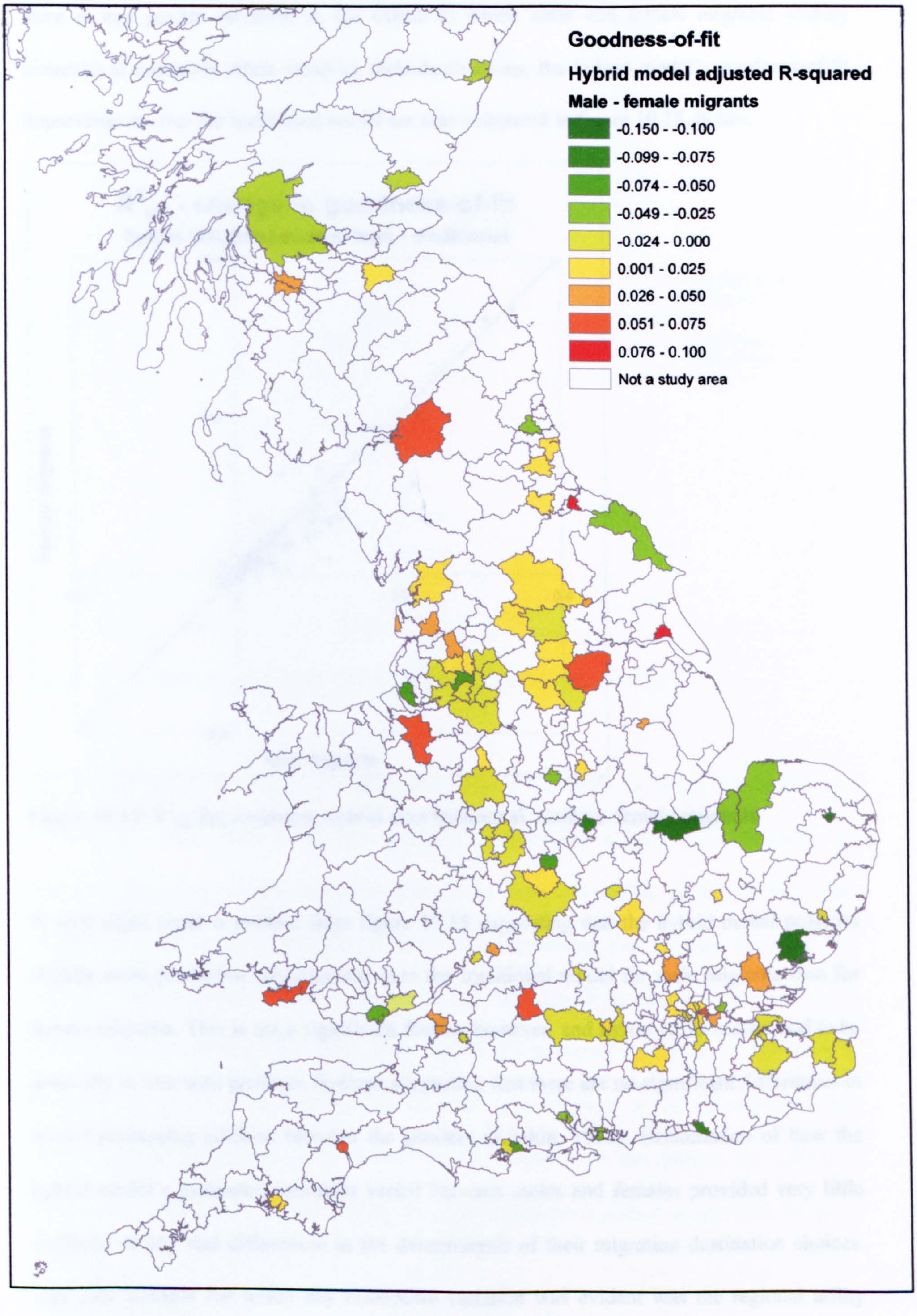


Figure 10.14: Hybrid model R^2_{adj}, male vs. female migrants.

It can be seen from figure 10.18 that whilst there is no systematic pattern in the goodness-of-fit between the genders, there is significant variation between origins in terms of which gender the model predicts migration more accurately for. Map 10.3, below, presents a spatial view of the difference in goodness-of-fit between males and females.



Map 10.3: Gender variation in goodness-of-fit of hybrid weighted nested logit model.

Again, no clear systematic pattern is evident from map 10.3. In order to investigate whether there is any gender variation in the extent to which male and female migrants employ hierarchical processes when selecting their destinations, the hybrid model's goodness-of-fit improvements over the traditional model are also compared in figure 10.15, below.

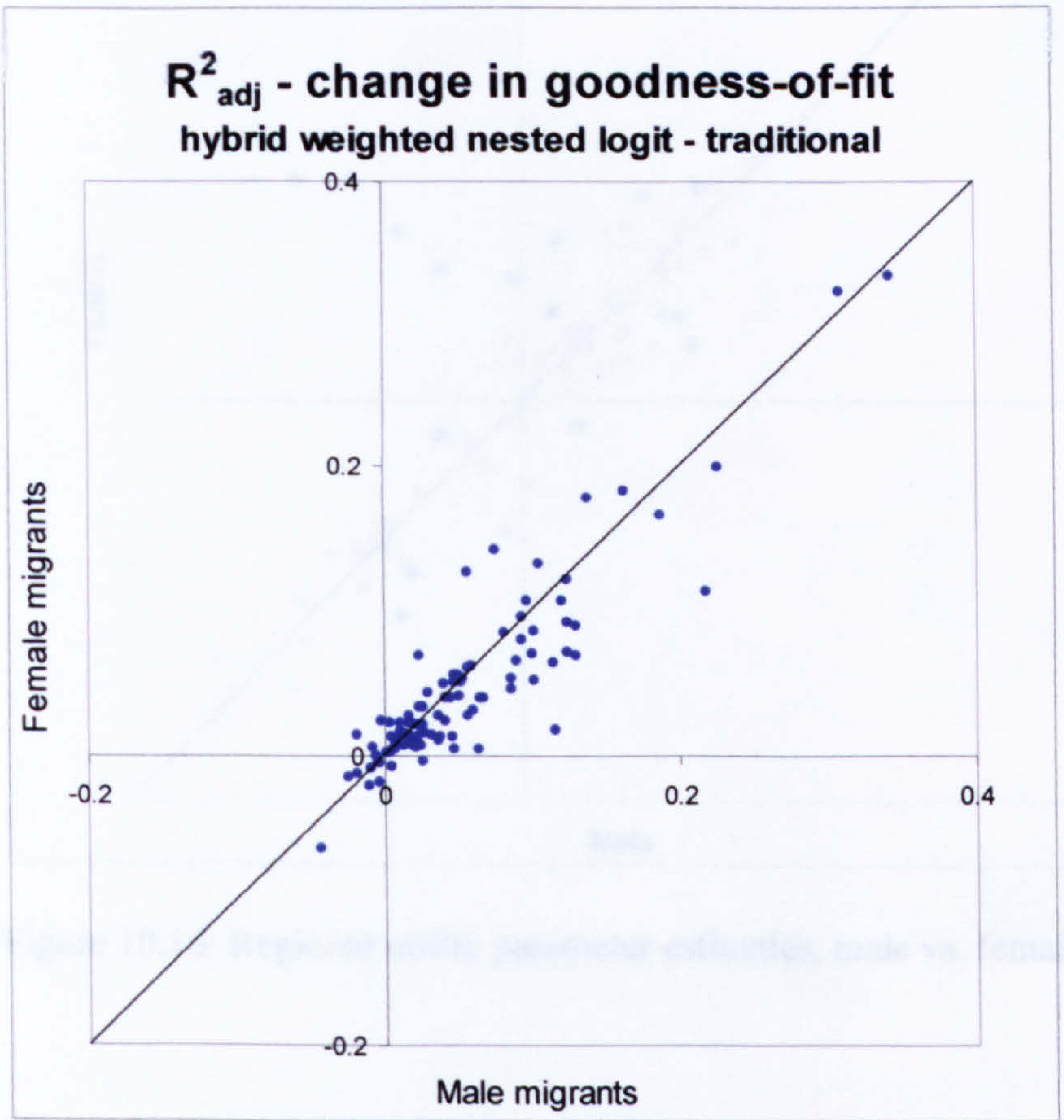


Figure 10.15: R^2_{adj} improvement, hybrid over traditional, male vs. female migrants.

A very slight trend is evident from figure 10.15 suggesting that the hybrid model provides slightly more predictive improvement over the traditional model for male migrants than for female migrants. This is not a significant finding however, and figure 10.15 can be said to be generally in line with previous findings suggesting that there are no significant differences in spatial processing abilities between the genders (Kitchin, 1996). Examination of how the hybrid model's parameter estimates varied between males and females provided very little evidence of any real differences in the determinants of their migration destination choices. The only variable for which any noticeable variation was evident was the regional utility variable – these differences are plotted in figure 10.16, below.

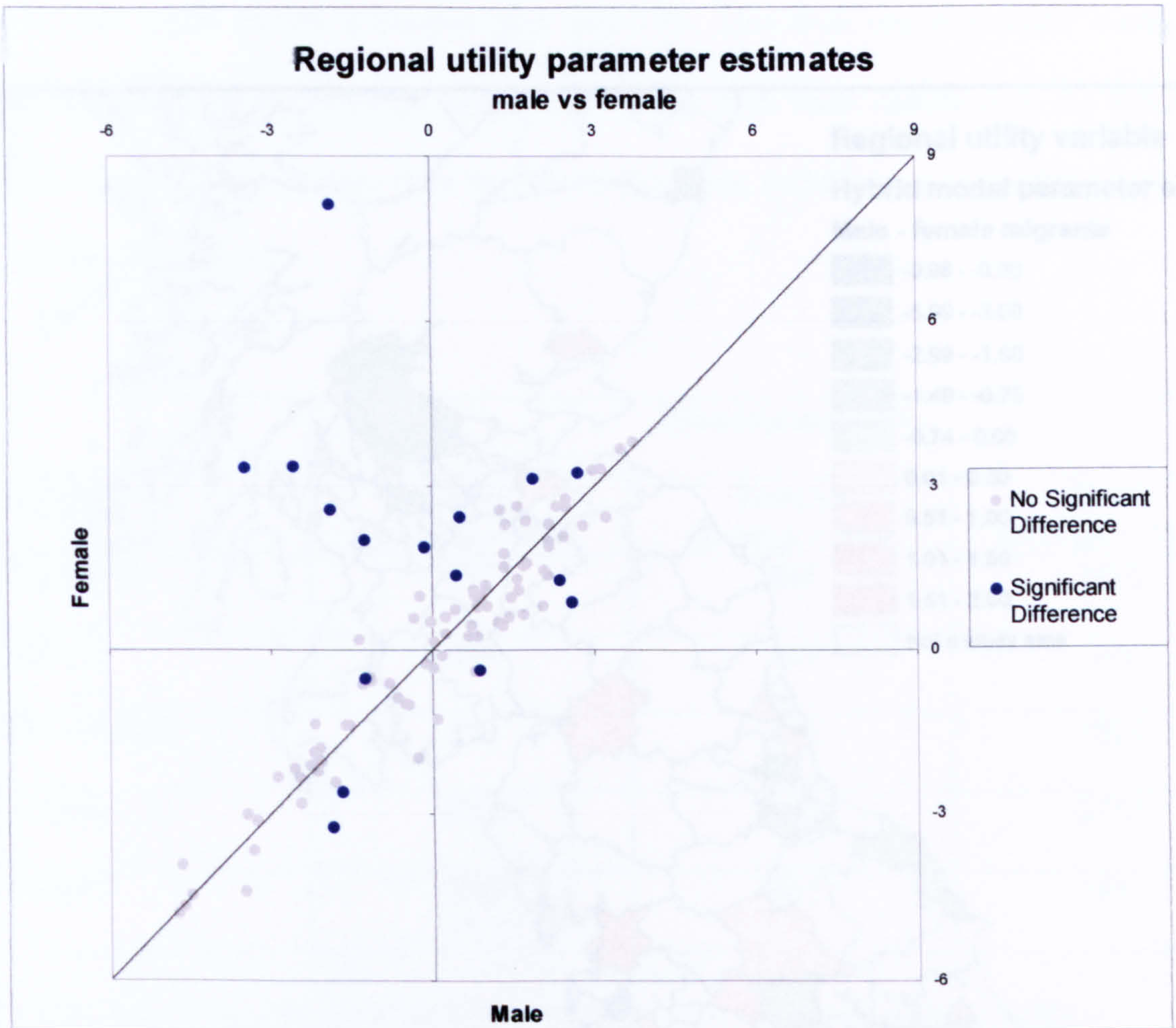
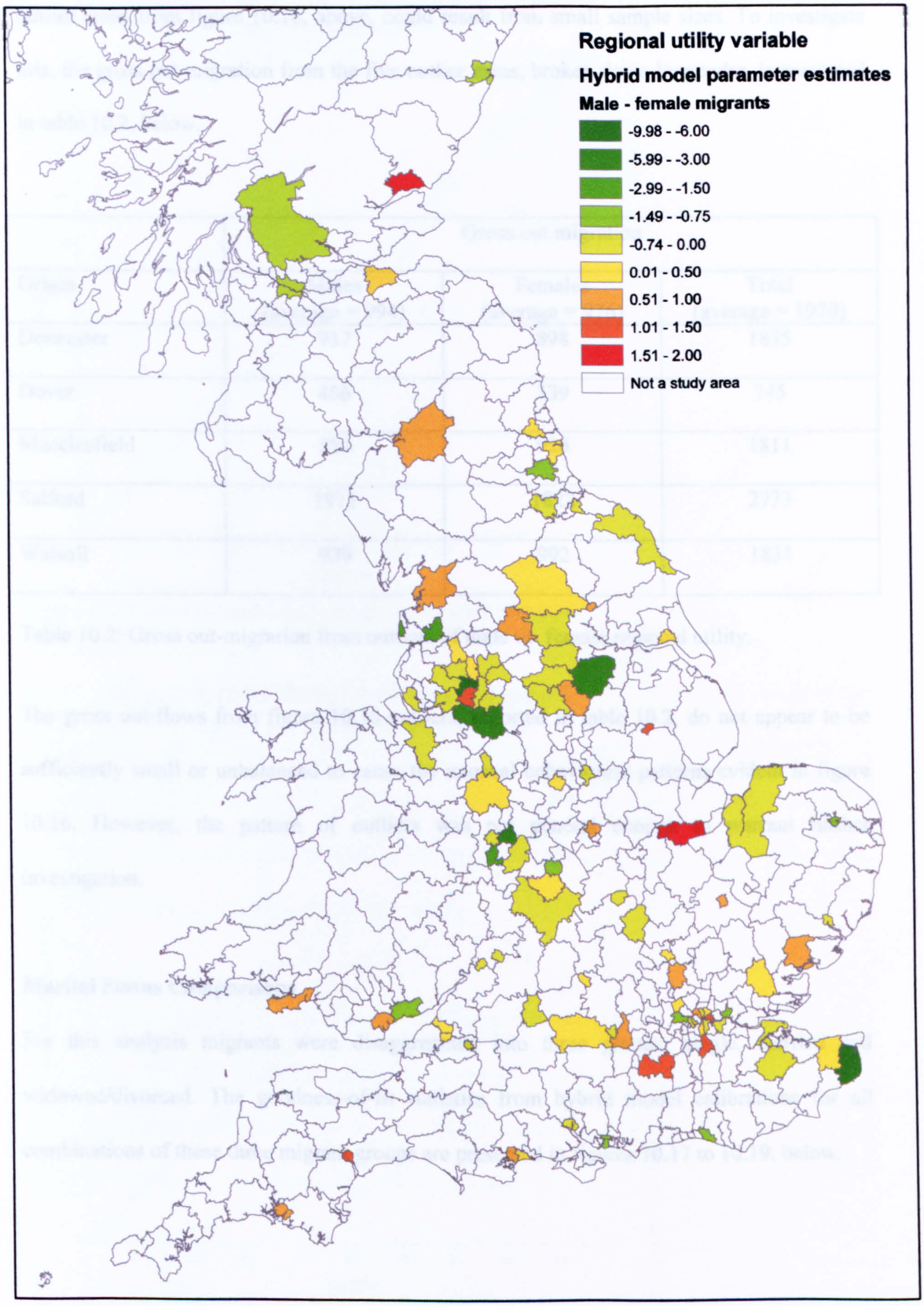


Figure 10.16: Regional utility parameter estimates, male vs. female migrants.

The majority of the parameter estimate pairs included on figure 10.16 are not significantly different from one another, but there are an interesting set of outliers in the upper left quadrant. These indicate that female migrants from these five origins are significantly more attracted to destinations with higher weighted regional utilities, than are male migrants. The regional utility parameter estimates for these five origins are all significantly different from 0 (at 95% level) as well as from each other. Just five exceptional origins is not sufficient to make any generalizations about gender variation in perception or use of regional utility variation. However, in case there is any spatial component to this pattern, the difference between male and female regional utility parameter estimates is plotted in map 10.4, below.



Map 10.4: Regional utility parameter estimates, male vs. female migrants.

No clear spatial pattern is evident from map 10.4. This leads one to consider whether the outlier areas from figure 10.16, above, could result from small sample sizes. To investigate this, the gross out-migration from the five outlier areas, broken down by gender, is presented in table 10.2, below.

	Gross out migration		
Origin	Males (average = 994)	Females (average = 976)	Total (average = 1970)
Doncaster	937	898	1835
Dover	406	339	745
Macclesfield	885	926	1811
Salford	1911	1822	2733
Walsall	939	992	1831

Table 10.2: Gross out-migration from outliers of male vs. female regional utility.

The gross out-flows from figure 10.16 outliers, reported in table 10.2, do not appear to be sufficiently small or unbalanced to cause the unusual behavioural patterns evident in figure 10.16. However, the pattern of outliers was not general enough to warrant further investigation.

Marital Status Comparisons

For this analysis migrants were disaggregated into three groups: single, married and widowed/divorced. The goodness-of-fit statistics from hybrid model calibrations for all combinations of these three migrant groups are presented in figures 10.17 to 10.19, below.

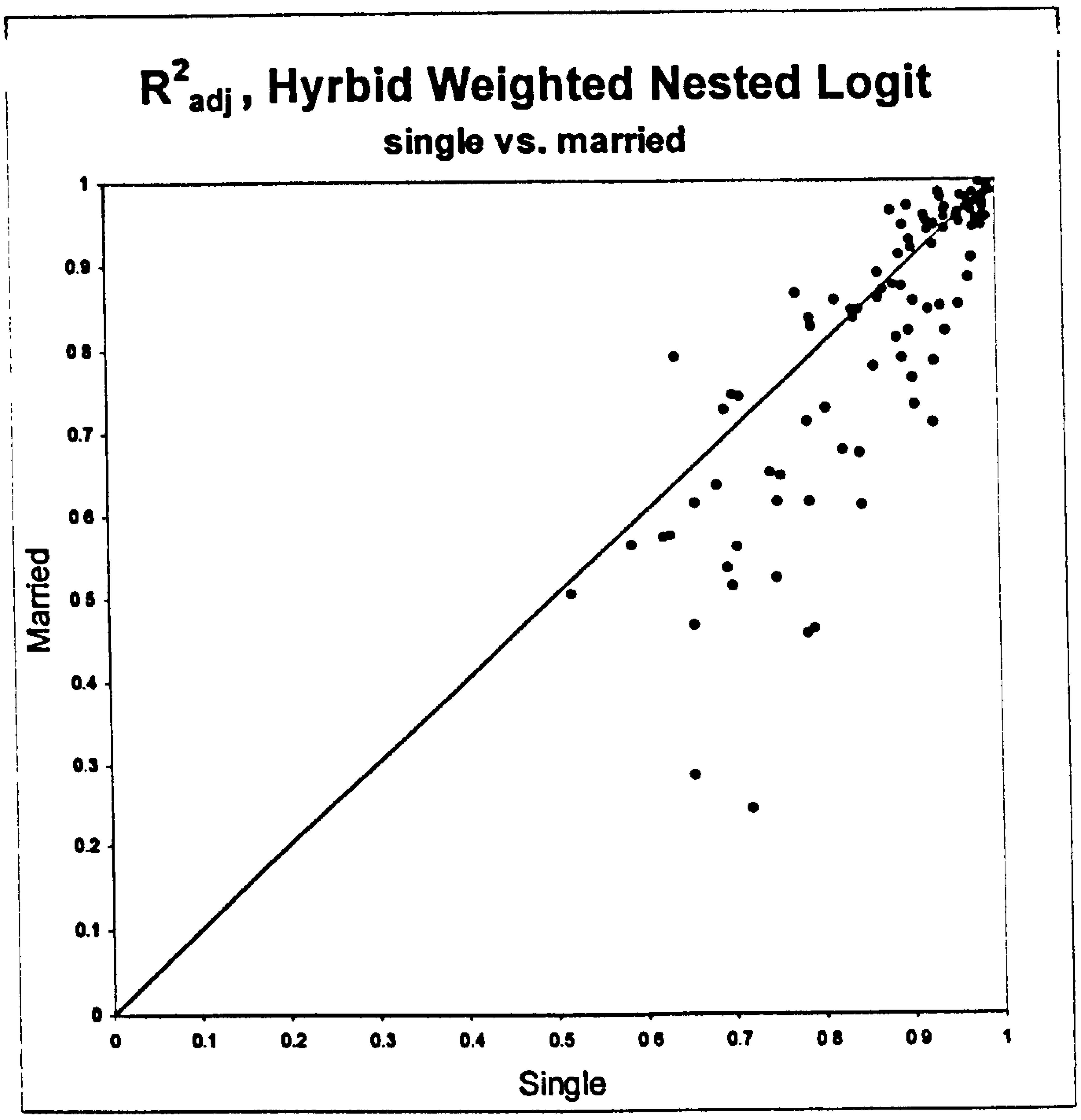


Figure 10.17: Hybrid model R^2_{adj}, single vs. married migrants.

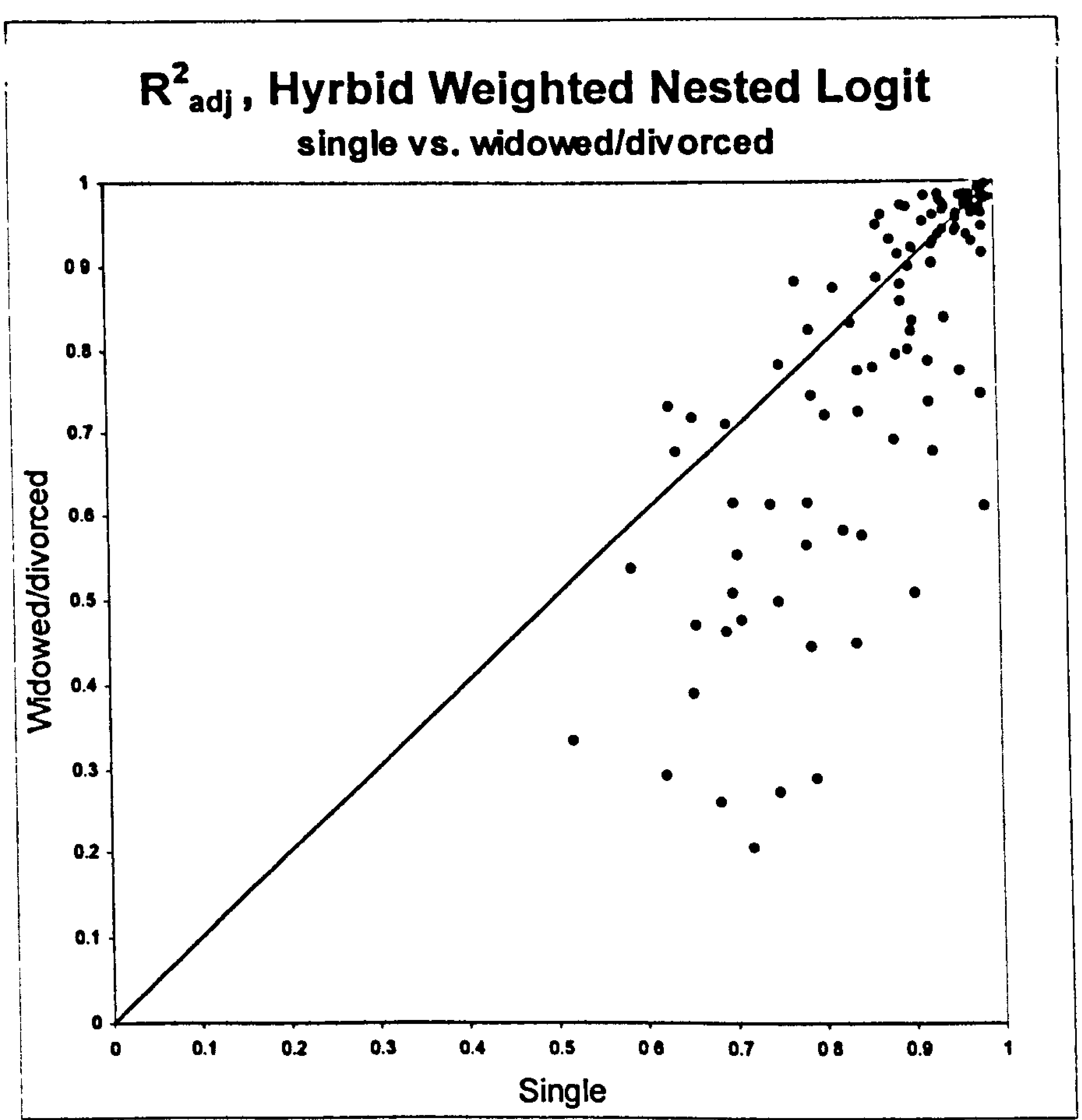


Figure 10.18: Hybrid model R^2_{adj}, single vs. widowed/divorced migrants.

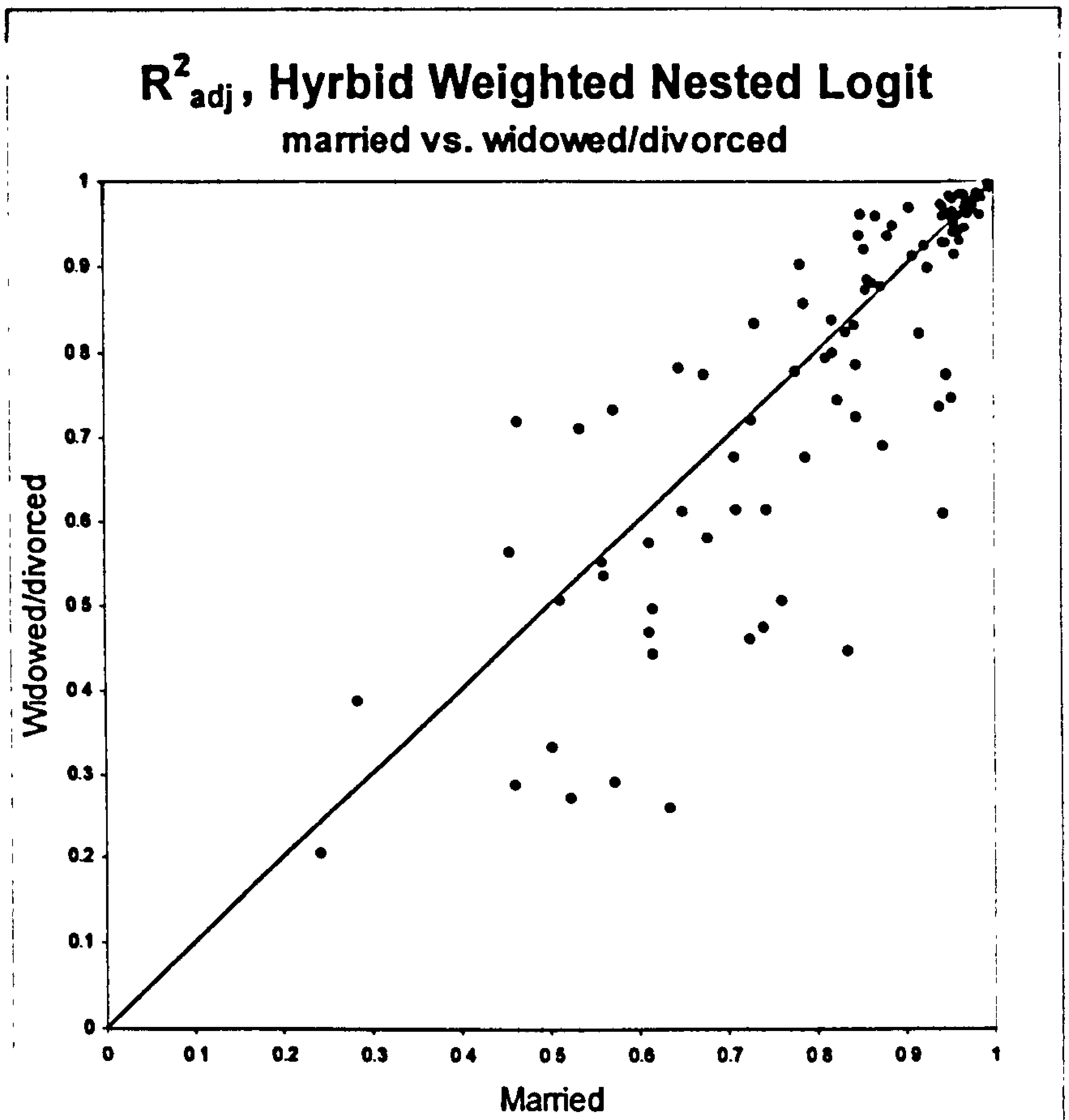


Figure 10.19: Hybrid model R^2_{adj}, married vs. widowed/divorced migrants.

It can be seen from figures 10.17 to 10.19 that when comparing the goodness-of-fit of the hybrid model for migrants of differing marital status, as when comparing behaviour of migrants of differing ages, the R^2_{adj} statistics produced by the model are affected heavily by the sample size against which they are calibrated, such that no meaningful direct comparison can be made between the R^2_{adj} values themselves.

Thus, as with the age group analysis, we investigate here whether there is any marital status variation in 'how hierarchically' migrants select their destinations, by comparing the hybrid model's goodness-of-fit improvement over the traditional model for the three marital status migrant sub-groups – see figures 10.20 to 10.22, below.

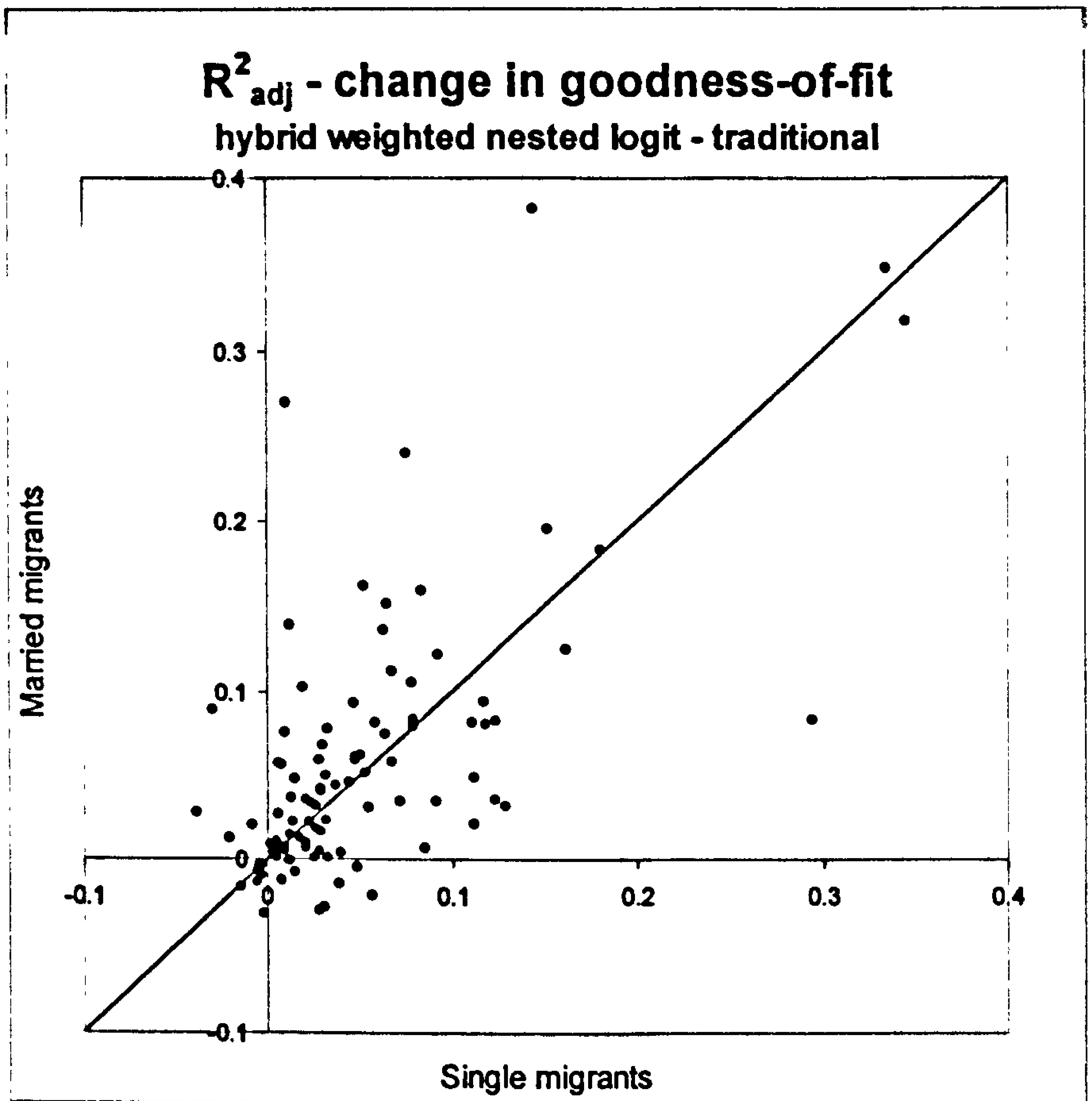


Figure 10.20: R^2_{adj} change, hybrid-traditional, single vs. married migrants.

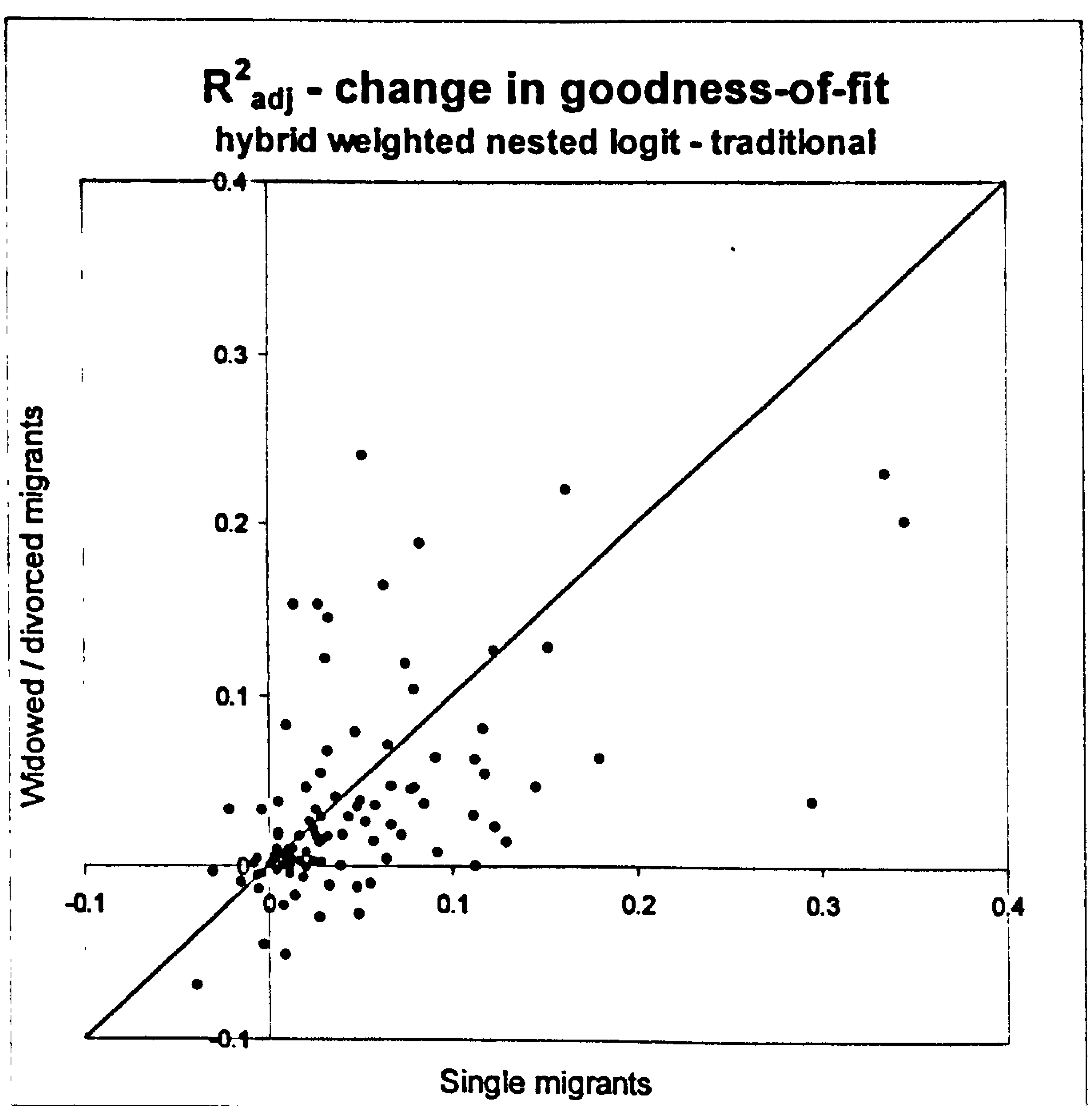


Figure 10.21: R^2_{adj} change, hybrid-traditional, single vs. widowed/divorced migrants.

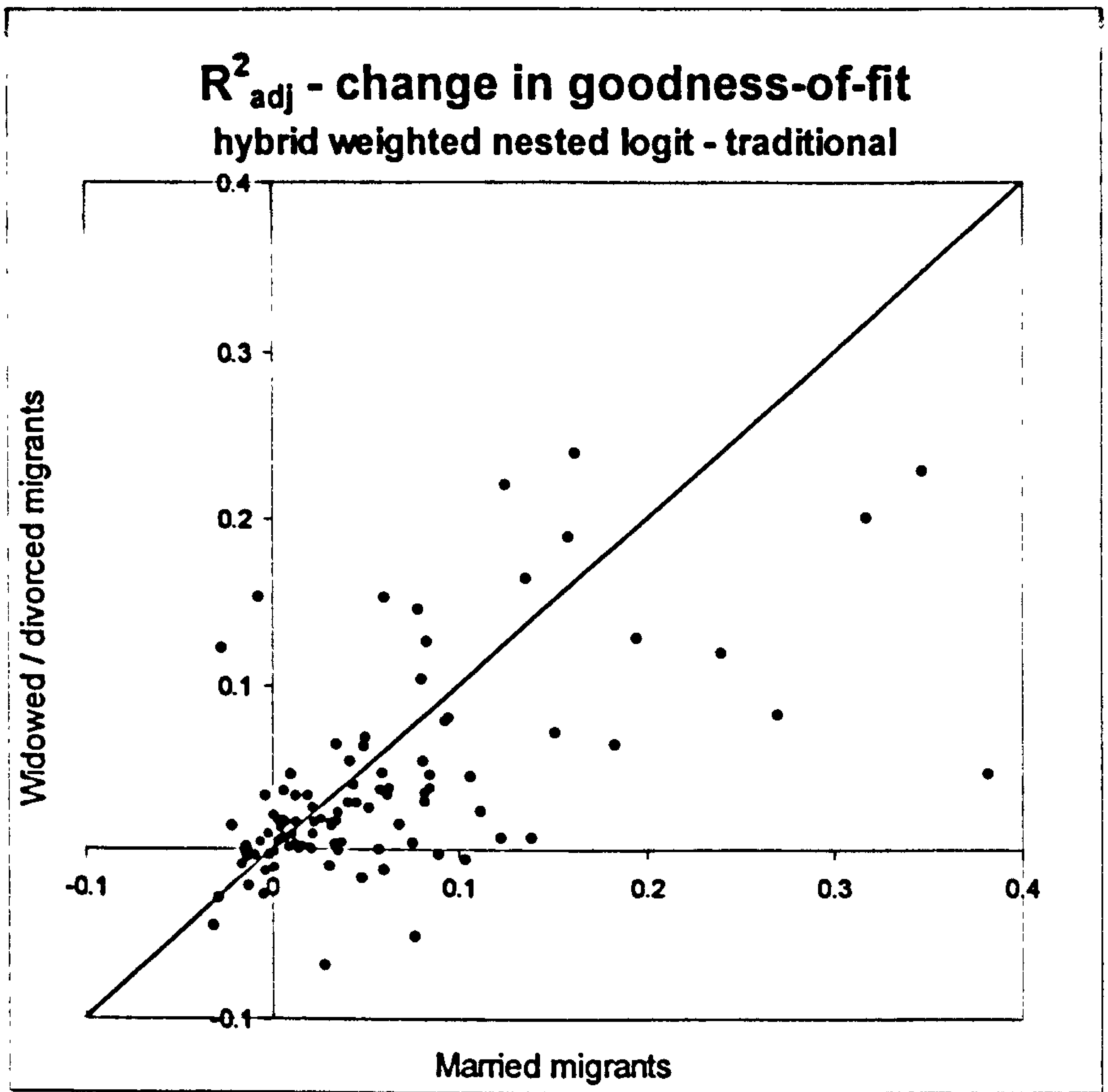


Figure 10.22: R^2_{adj} change, hybrid-traditional, married vs. widowed/divorced migrants

The comparison of the hybrid models improvements over the traditional model, for the different marital status migrant subgroups, presented in figures 10.20 to 10.22, provides no evidence of any difference in the extent of hierarchical processing by the different marital status groups.

However, examination of the variables' parameter estimates for the different marital status migrant subgroups does provide some interesting results.

Single vs. married migrants

Figures 10.23 to 10.25, below, compare single and married migrants' hybrid model parameter estimates for the social class, tenure and unemployment variables, respectively.

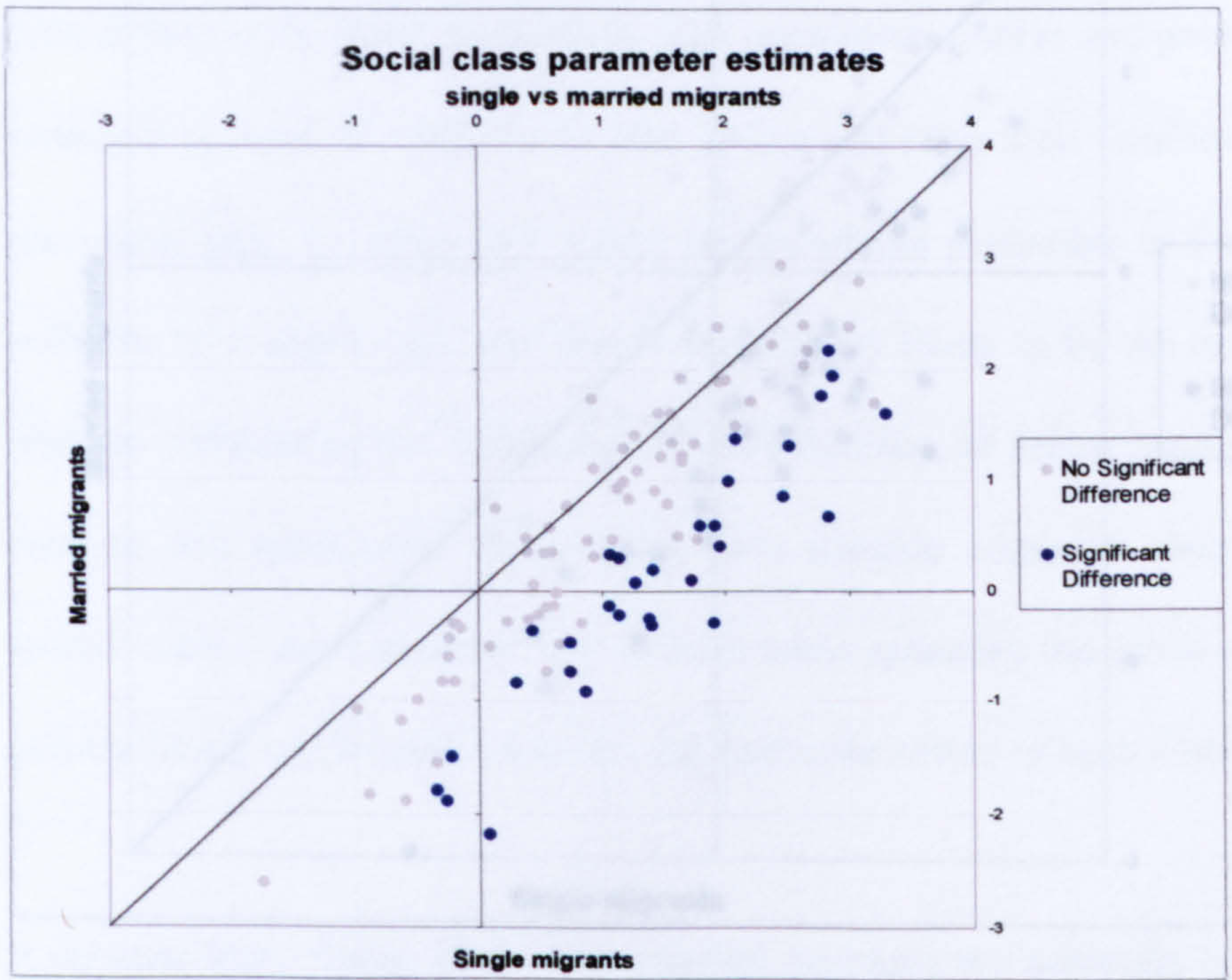


Figure 10.23: Social class parameter estimates, single vs. married migrants.

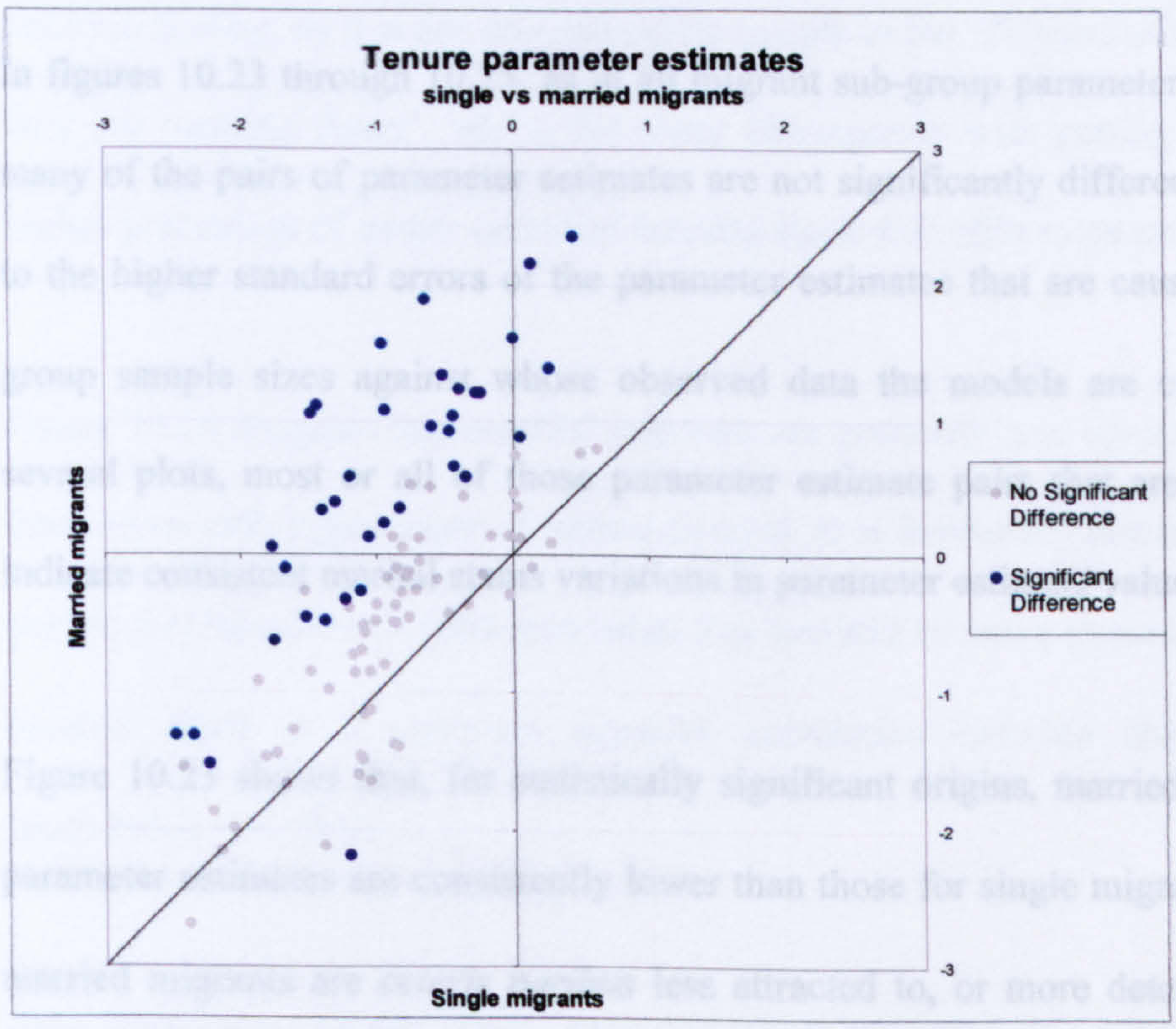


Figure 10.24: Tenure parameter estimates, single vs. married migrants.

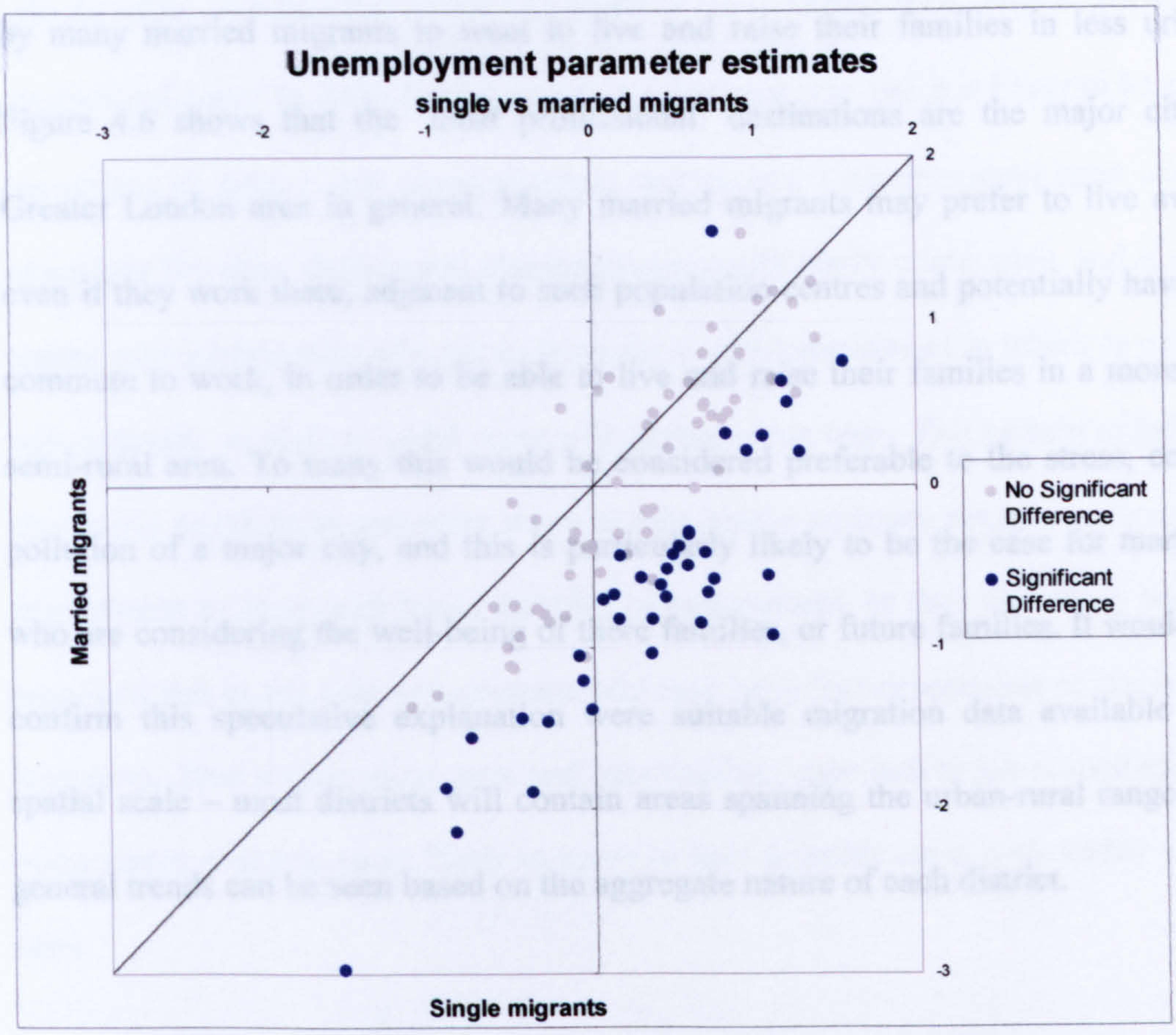


Figure 10.25: Unemployment parameter estimates, single vs. married migrants.

In figures 10.23 through 10.25, as in all migrant sub-group parameter estimate comparisons, many of the pairs of parameter estimates are not significantly different from each other, due to the higher standard errors of the parameter estimates that are caused by the smaller sub-group sample sizes against whose observed data the models are calibrated. However, in several plots, most or all of those parameter estimate pairs that are significantly different indicate consistent marital status variations in parameter estimate values.

Figure 10.23 shows that, for statistically significant origins, married migrants' social class parameter estimates are consistently lower than those for single migrants. This indicates that married migrants are *ceteris paribus* less attracted to, or more deterred from areas with a more professional social class structure, (recall that the social class variable is the percentage of the workforce in professional or senior managerial positions). This could reflect a desire by many married migrants to want to live and raise their families in less urban settings. Figure 4.6 shows that the 'most professional' destinations are the major cities, and the Greater London area in general. Many married migrants may prefer to live away from or, even if they work there, adjacent to such population centres and potentially have to endure a commute to work, in order to be able to live and raise their families in a more suburban or semi-rural area. To many this would be considered preferable to the stress, congestion and pollution of a major city, and this is particularly likely to be the case for married migrants who are considering the well-being of there families, or future families. It would be easier to confirm this speculative explanation were suitable migration data available at a smaller spatial scale – most districts will contain areas spanning the urban-rural range so only very general trends can be seen based on the aggregate nature of each district.

It appears from figure 10.24 that married migrants are generally more attracted to or less deterred from areas with a higher percentage of owner occupied housing stock. This is an intuitive finding, as it is not uncommon for people to put off purchasing their first home until they are 'settling down', which for many corresponds with getting married. Areas with a higher percentage of owner occupied housing stock will offer more property for purchase.

Figure 10.25 suggests that married migrants are generally less attracted to or more deterred from areas with higher rates of unemployment. It is interesting that married migrants should exhibit this behaviour, whilst also being less inclined to move to areas of higher social class, because there is a moderate negative correlation between these two characteristics (correlation = -0.581).

This behaviour could result from married migrants considering areas with lower unemployment to be better environments in which to settle down and potentially raise a family. Lower unemployment rates could be perceived as meaning an area both has a lot of jobs to offer, and also has lower competition for available jobs. This is not necessarily an accurate inference to make from a low unemployment rate, but it is likely to be w widespread perception of what a low unemployment rate means. Unemployment rates can also act as surrogates for other detrimental characteristics, particularly crime rates, and it is intuitive to expect married migrants who are looking for a safe environment in which to settle down and raise a family, to want to avoid areas with higher crime rates. This pattern of behaviour could also partly result from generally more mobile single migrants being more prepared to seek employment in areas with lower stability of employment, as they will likely have less socio-economic ties to the area and consequently may have less expectation of staying in that area long-term. Thus a more short-term and opportunistic approach to job-related migration might make single migrants more likely to move to less desirable areas with higher unemployment rates.

Single vs. widowed/divorced migrants

Several interesting patterns are evident when comparing the hybrid model's parameter estimates for single and widowed/divorced migrant sub groups. Figures 10.26 through 10.29, below, show the most marked trends.

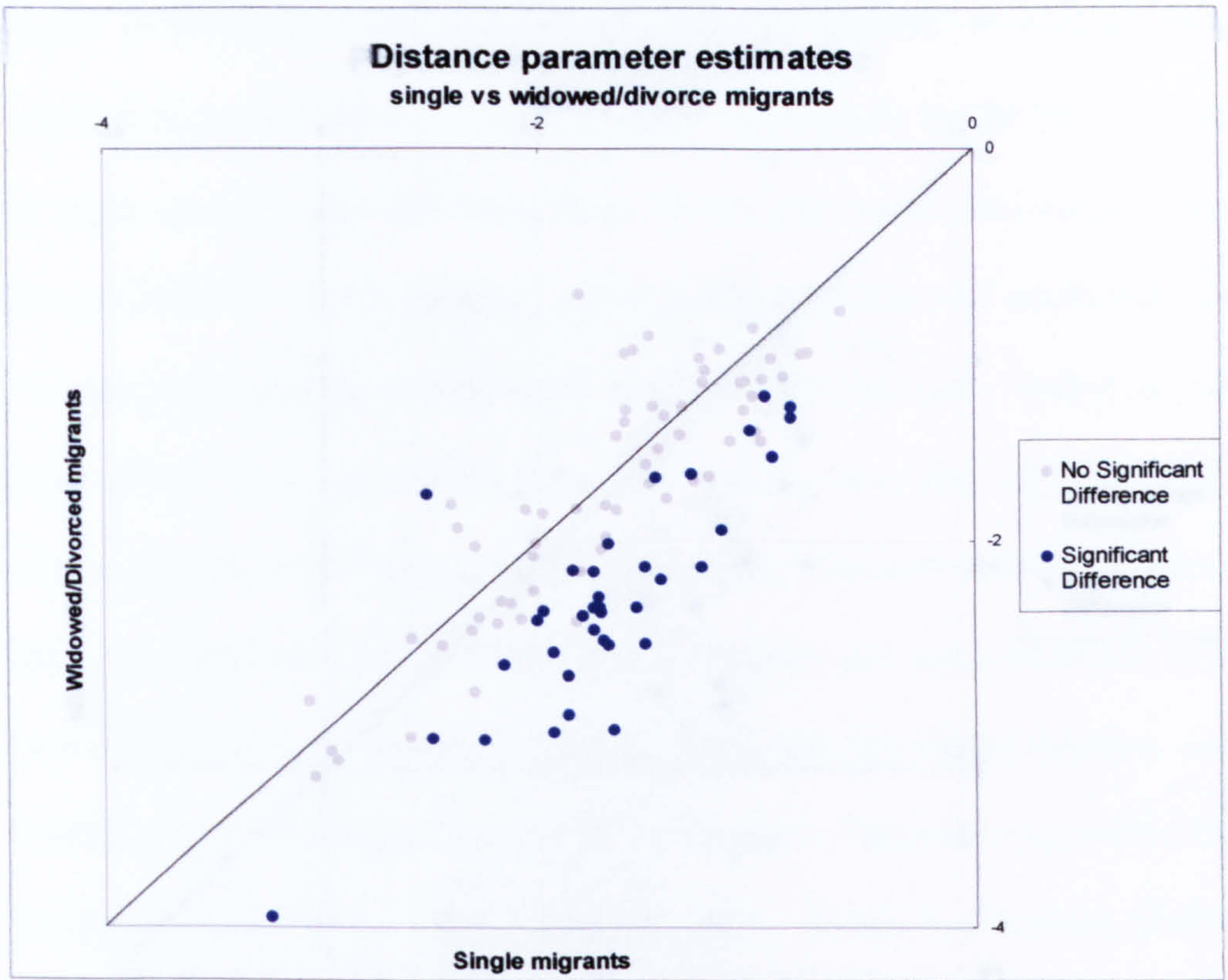


Figure 10.26: Distance parameter estimates, single vs. widowed/divorced migrants.

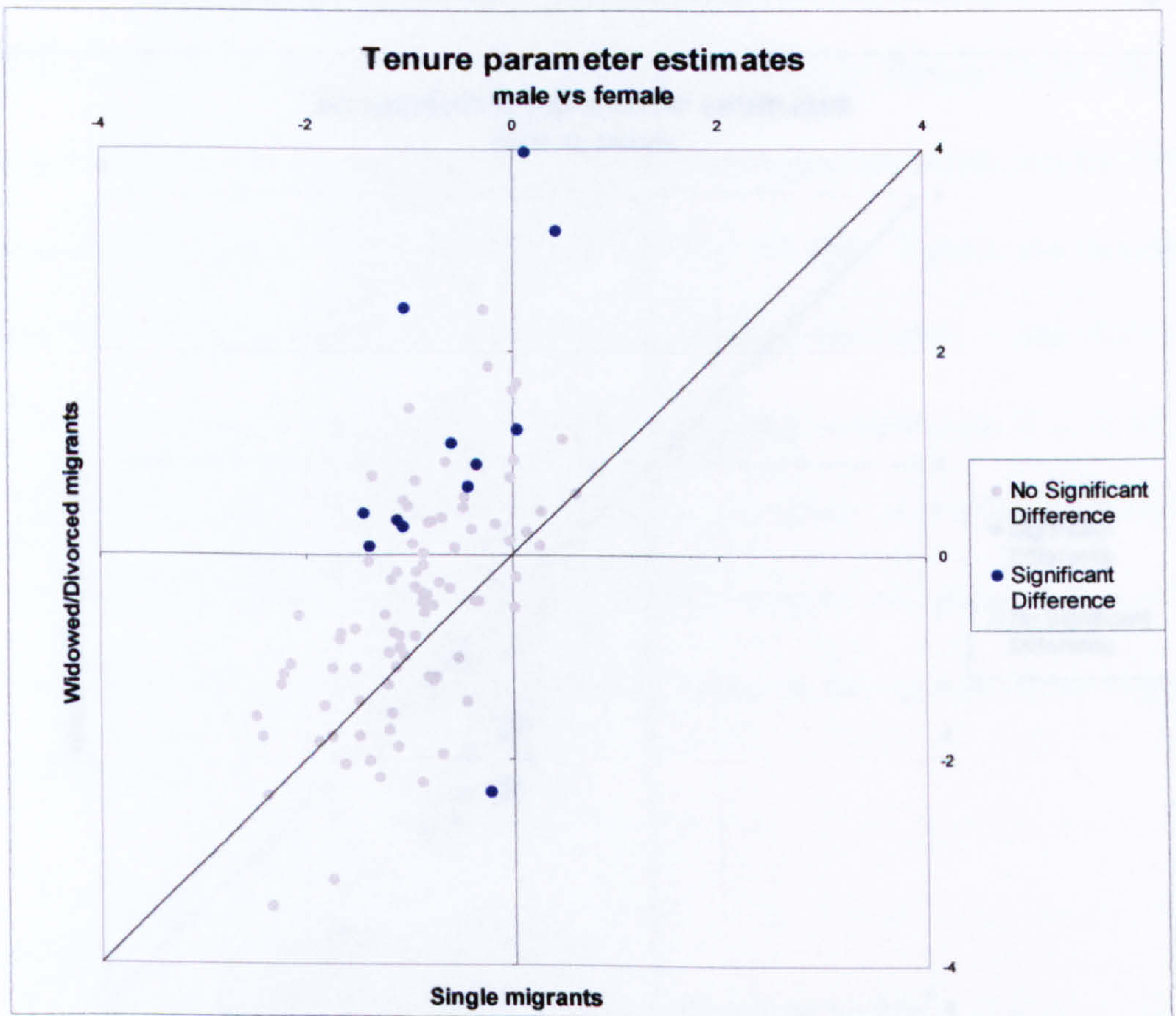


Figure 10.27: Tenure parameter estimates, single vs. widowed/divorced migrants.

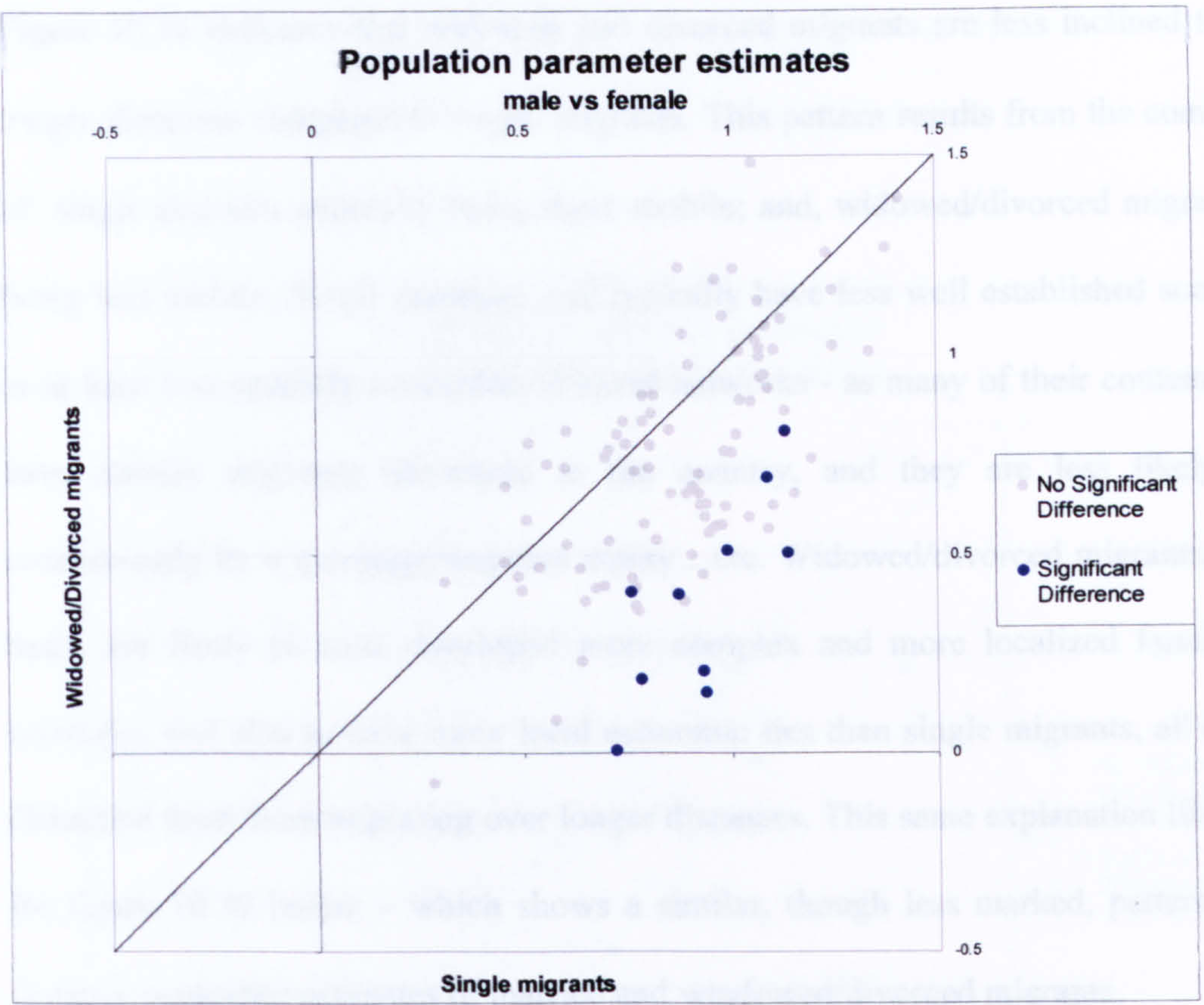


Figure 10.28: Population parameter estimates, single vs. widowed/divorced migrants.

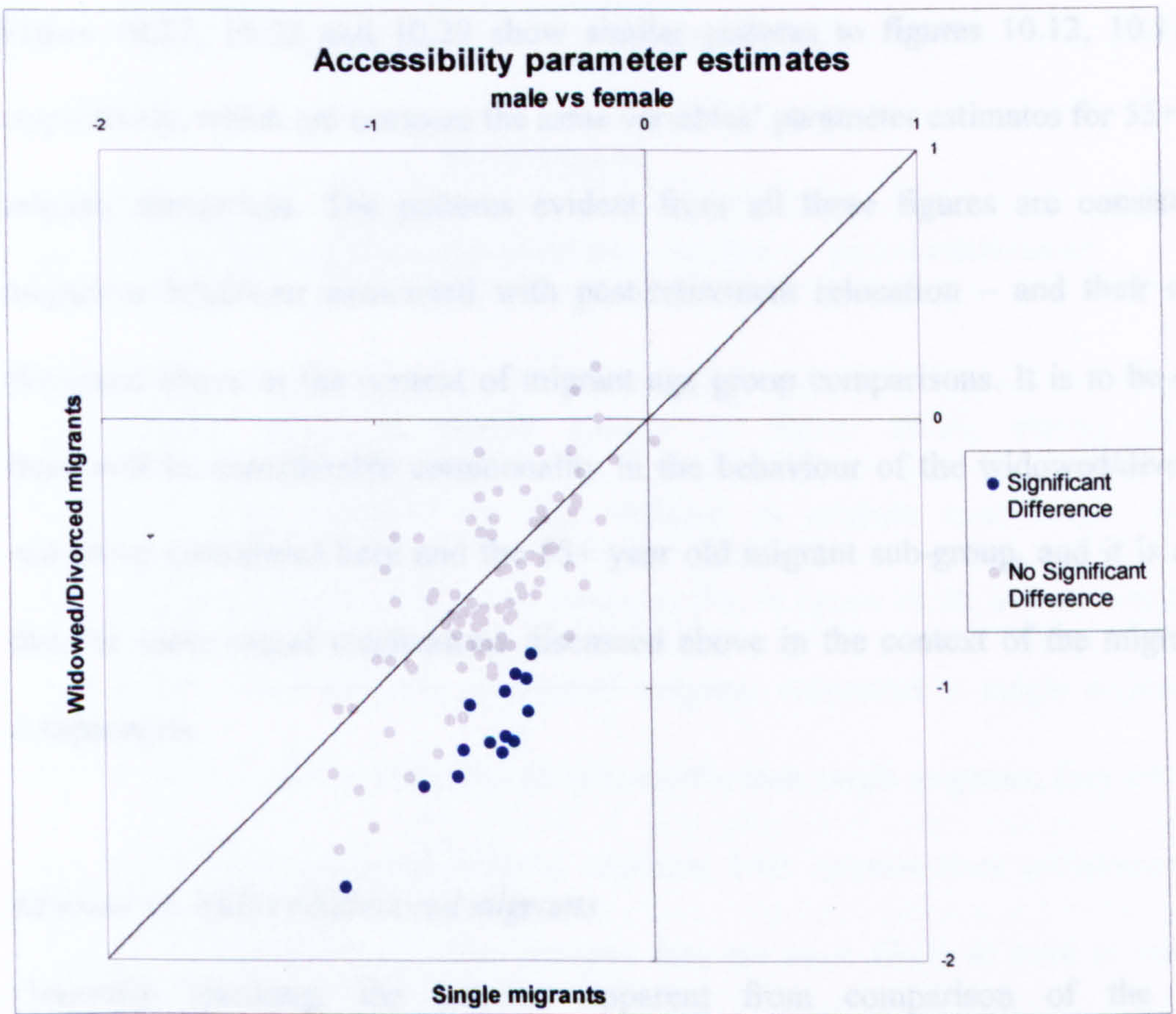


Figure 10.29: Accessibility parameter estimates, single vs. widowed/divorced.

Figure 10.26 indicates that widowed and divorced migrants are less inclined to move over longer distances compared to single migrants. This pattern results from the combined effects of: single migrants generally being more mobile; and, widowed/divorced migrants generally being less mobile. Single migrants will typically have less well established social networks, or at least less spatially concentrated social networks - as many of their contemporaries may have already migrated elsewhere in the country, and they are less likely to be tied economically by a mortgage/negative equity...etc. Widowed/divorced migrants, on the other hand, are likely to have developed more complex and more localized family and social networks, and also to have more local economic ties than single migrants, all of which will disincline them from migrating over longer distances. This same explanation likely holds true for figure 10.30 below – which shows a similar, though less marked, pattern between the distance parameter estimates of married and windowed/divorced migrants.

Figure 10.27, 10.28 and 10.29 show similar patterns to figures 10.12, 10.11 and 10.13, respectively, which are compare the same variables' parameter estimates for 55+ and younger migrant sub-groups. The patterns evident from all these figures are consistent with the migration behaviour associated with post-retirement relocation – and their causality was discussed above in the context of migrant age group comparisons. It is to be expected that there will be considerable commonality in the behaviour of the widowed/divorced migrant sub-group considered here and the 55+ year old migrant sub-group, and it is assumed here that the same causal mechanisms discussed above in the context of the migrant age-group comparisons.

Married vs. widowed/divorced migrants

Generally speaking, the patterns apparent from comparison of the married and widowed/divorced migrants reinforce those seen when comparing single and widowed/divorced migrants' behaviour. Figures 10.30 and 10.31, below, present the strongest parameter estimate patterns resulting from comparisons of married and widowed/divorced migrants.

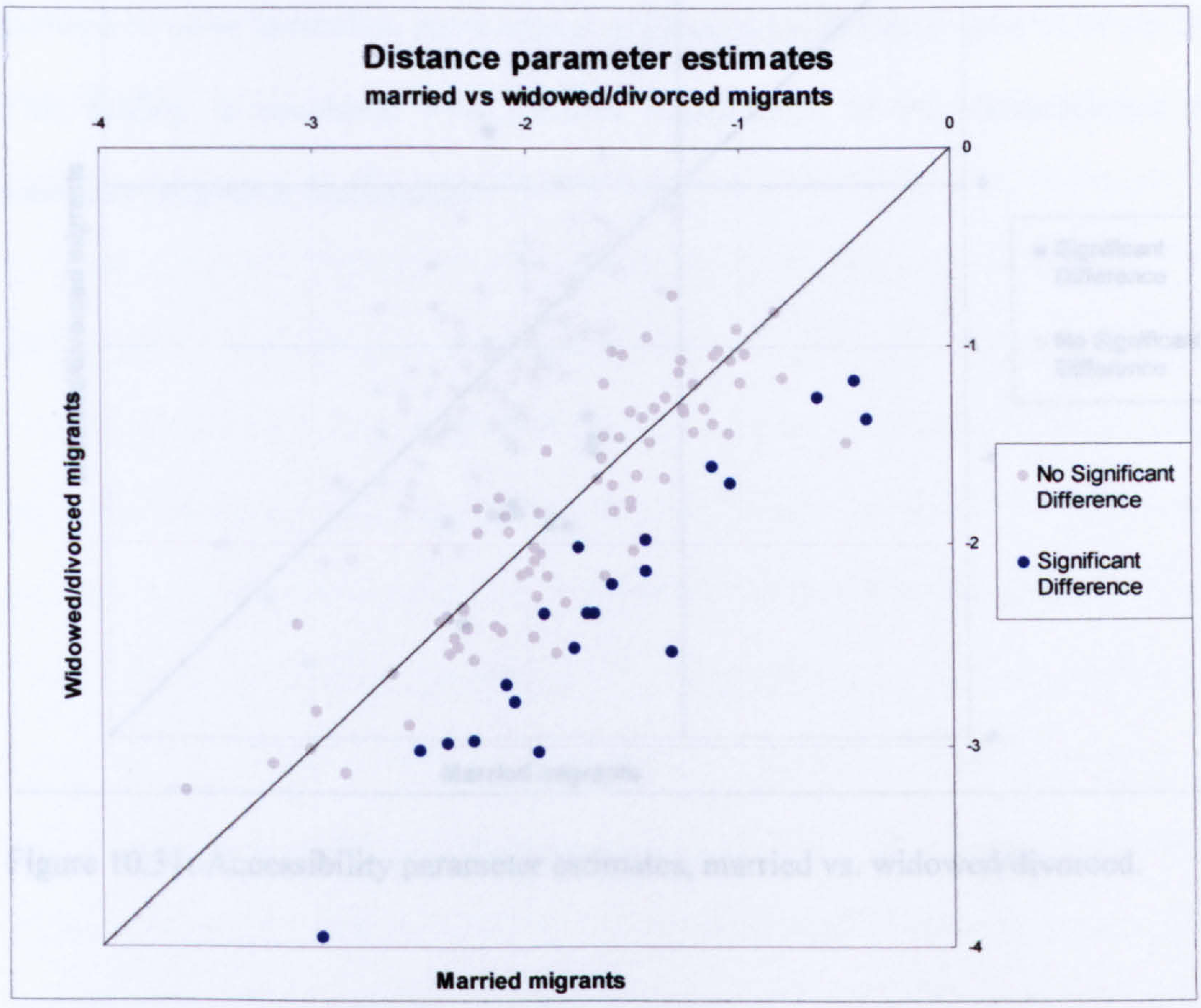


Figure 10.30: Distance parameter estimates, married vs. widowed/divorced.

Figure 10.30, shows a similar pattern to figure 10.26, above, indicating that widowed/divorced migrants are less inclined to migrate over longer distances. The relationship in figure 10.30 is not as strong as that in figure 10.26, as one would expect given the generally lower mobility of married migrants compared to single migrants. However, whilst married migrants will often be less mobile than single migrants, they are still generally more mobile than widowed/divorced migrants, both because their social/economic ties are not as well established, and also because they are more likely to have to make job-related migrations/relocations that could be over longer distances.

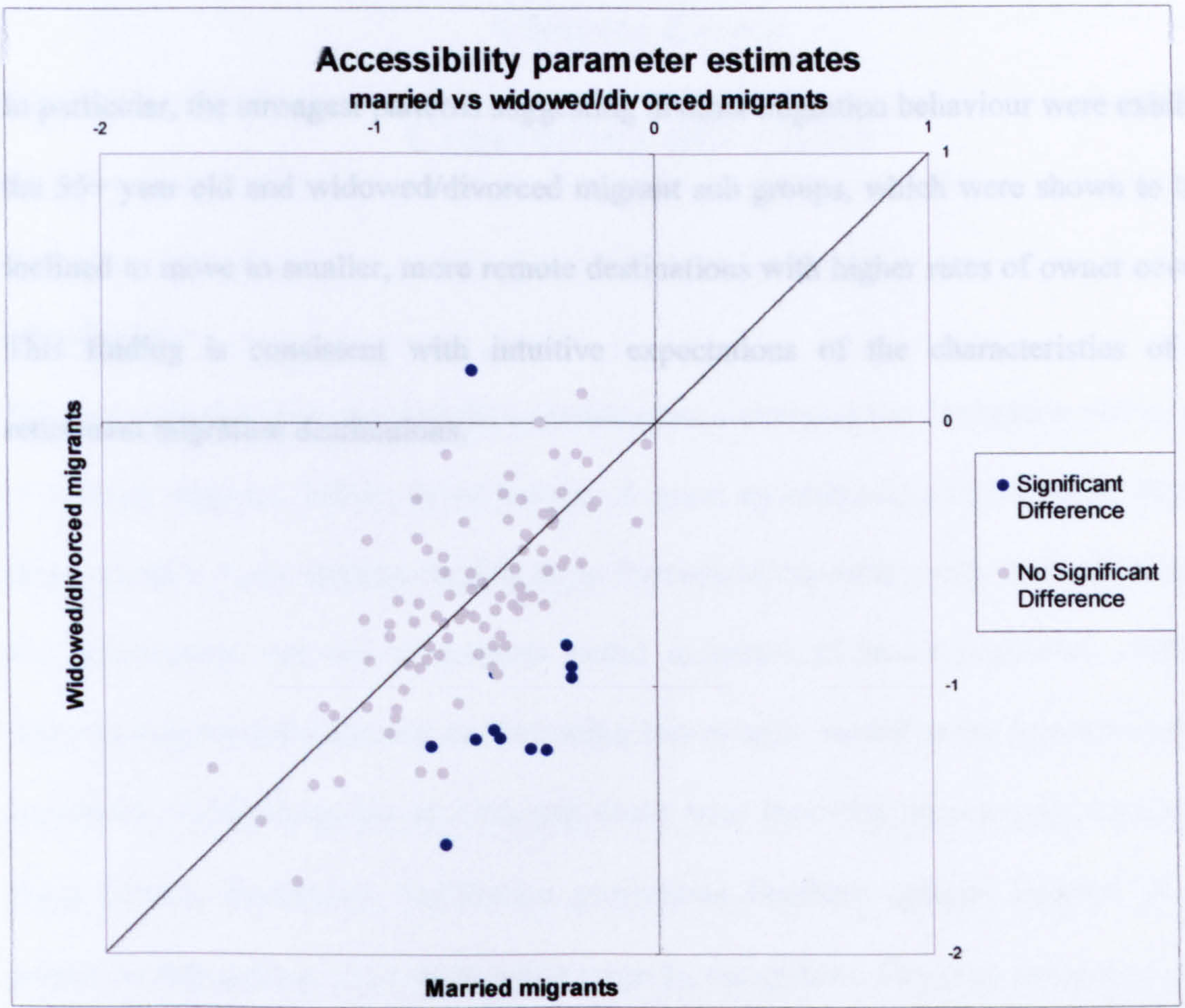


Figure 10.31: Accessibility parameter estimates, married vs. widowed/divorced.

As discussed above when comparing single and widowed/divorced migrants, the latter groups' retirees are generally more likely to select destinations with lower accessibility. Conversely, married migrants will often feel pressure to migrate to destinations with moderate to high accessibility as those are likely to be the areas, close to major population centres, where the majority of employment opportunities will be found.

Summary

Whilst this examination of variation in migration destination choice behaviour between migrant subgroups has not been able to draw meaningful conclusions about goodness-of-fit of the models to the different migrant subgroups, (because of the varying sample sizes of the subgroups), many interesting patterns were uncovered when parameter estimates were compared between subgroups.

In particular, the strongest patterns suggesting distinct migration behaviour were exhibited by the 55+ year old and widowed/divorced migrant sub groups, which were shown to be more inclined to move to smaller, more remote destinations with higher rates of owner occupancy. This finding is consistent with intuitive expectations of the characteristics of typical retirement migration destinations.

Chapter Eleven

Concluding discussion

Introduction

This research examines the cognitive mechanisms underlying the destination choice process of internal migrants within Great Britain. A good understanding of migration destination choice enables more accurate models to be formulated that both public and private agencies and corporations can use to generate better estimates of future population distribution. Accurate migration predictions are becoming increasingly central to the forecasting of future population distributions due to birth and death rates becoming increasingly aspatial across Great Britain. Population distribution projections facilitate optimal location of various amenities: fire stations, civil engineering projects, out-of-town shopping complexes...etc. not only with respect to current population patterns but also anticipating where people will be living in the future. This can save lives and money, make businesses more profitable and increase a population's retail and recreational options, amongst many other possible benefits.

Motivation for this research

The catalyst for this research was the emergence of a new generation of spatial interaction models that address some of the shortcomings of traditional flat-processing approach to spatial interaction modelling. This next generation of spatial interaction models take account of the spatial structure of the choice set. Application of these emerging models, and other novel models derived from them, to the analysis of internal migration within Great Britain provides a useful way to compare and contrast the performance of the models and draw some conclusions from that about the psychological nature of the decision process underlying migration destination choice.

Research question

Specifically, this research attempts to demonstrate that migration destination choices reflect a hierarchical decision-making process in which migrants group potential destinations based on the spatial structure present in the destination choice set and then evaluate some characteristics at a group level before selecting a specific destination from a preferred group. It is argued that this is a much more reasonable approach than the traditional model's 'flat-processing' assumption that each migrant has full information about and gives equal consideration to every possible destination.

Epistemology

In order to provide results that can be generalized in a way that is useful to public and private organizations with an interest in future population patterns, a quantitative model-based approach is adopted that follows the traditional deterministic positivist epistemology that has become pervasive in quantitative human geography.

Specifically, a data-driven, revealed preference approach to the analysis of migration destination choice has been employed based on the calibration and comparison of a number of migration models some of which have been derived from hierarchical principles. This approach was considered to be preferable to qualitative methods not only because its results can be scaled to provide more broadly applicable conclusions and recommendations, but also because survey or interview based approaches are less appropriate when considering a phenomenon such as migration behaviour, some aspects of which are very likely subconscious, or at least less-than-conscious. This may well be particularly true of the hierarchical aspects of destination choice – migrants probably do not make a conscious distinction between the considerations of regional-level versus destination-specific characteristics.

Whilst this research has employed quantitative analytical techniques to the investigation of migration destination choice, it should be stressed that the derivation of several of the hierarchical models applied here is based upon a consideration of individual-level decision-making processes. This contrasts with earlier application of statistical migration models, which were derived at a more aggregate-level, such as Ravenstein's 'social gravity' models, or which are based on entirely impersonal concepts, such as Wilson's entropy maximizing models. This consideration of individual level behavioural patterns can be said to add a humanist element to this analysis, and to bring aggregate level migration analysis closer to the grass-roots of individual-focused migration research that has traditionally been the domain of purely qualitative research methods.

Another noteworthy distinction between this analysis and the majority of earlier analysis is the focus on local model calibrations. This is in recognition of the fact that migration behaviour will inevitably vary between individual migrants. Whilst it is not possible, given the limitations of available data, or necessarily even desirable, in terms of the ability to draw widely applicable conclusions, to model migration at the individual-level, it is entirely achievable and useful to examine differences in migration behaviour between migrants leaving different origins, and that is the approach that has been followed throughout this research.

In general, origin-specific models have been calibrated against the observed migration of all migrants (aged 16+ years), but the origin-specific out-migration of several sub-groups have also been modelled, disaggregating migrants by: age, gender and marital status.

Methodology

The approach taken to addressing the central research question of this thesis was to apply a number of latest generation spatial interaction models to the analysis of migration destination choice, and from comparisons of their performance relative to each other and to a traditional non-hierarchical model, to draw conclusions about which models' underlying psychological assumptions are most applicable to the behaviour of actual internal migrants within Great Britain. All models were calibrated locally for each of 100 selected migration origins in order to capture local variation in migration destination choice behaviour. The same set of explanatory data describing potential migration destinations was used by all models, this included each destination's: population, distance from origin, social class structure (% of workforce in senior management or professional jobs), weighted average house price, tenure structure (% of housing stock that is owner occupied) and unemployment rate. Accessibility and/or regional utility variables are also included in the various hierarchical models – these variables are intended to capture the effects of the spatial structure in migrants' destination choice sets, and it is these variables that differentiate these next-generation hierarchical models from traditional aspatial discrete choice models of migration destination choice.

The performance of the various models was assessed in terms of goodness-of-fit, using the Adjusted R-squared and Akaike Information Criterion measures of model fit, and also through examination of error flow residuals for selected migration origins. The parameter estimates from the various models were also compared and any systematic differences investigated.

Four distinct hierarchical migration destination choice models were applied in addition to the traditional model, these were the: competing destinations, discrete nested, weighted nested logit and hybrid weighted nested logit models.

The competing destinations model captures spatial structure effects, and their effect upon migrants' destination choice behaviour, through the inclusion of an accessibility variable. This variable represents the likelihood of a destination being considered by any particular migrant as being contained within a larger cluster of destinations. This is important because of a documented psychophysical observation that people increasingly under-estimate cluster size as actual cluster size increases, or, in this migration context, as the number of competing destinations increases. If migration destinations are selected in a hierarchical way, with some characteristics being considered at a collective spatial scale of region or destination cluster, then one would expect that under-estimation of the membership of larger clusters would cause individual destinations within those clusters to be selected less often than their characteristics would otherwise warrant. The competing destinations model of migration destination choice can capture this spatial structure effect whereas the traditional model cannot.

The discrete nested logit model is a model that has been widely applied to the analysis of decisions such as brand choice or selection of mode of transportation. This model requires that the entire choice set be pre-allocated to mutually-exclusive categories, in the context of migration destination choice this equates to grouping all destinations into discrete regions, such that every district is contained in exactly one region. When a discrete nested logit model is calibrated this discrete regionalization of destinations is used to calculate each destination's discrete regional utility – this is simply the sum of the utilities of all destinations in that destination's region, so each district within the same region will have the same value for the regional utility variable. Statistically significant parameter estimates for the discrete regional utility variable indicate that migrants are examining characteristics at a regional level as well as at a destination level, which is clear evidence that migrants are employing partly hierarchical migration destination choice processes.

The weighted nested logit model operates in a similar manner to the discrete nested logit model but it is calibrated within the context of a probabilistic definition of regions - which is a 459-square matrix each cell of which represents the likelihood of a particular pair of destinations being cognized together in the same destination cluster by a migrant. A weighted regional utility variable is then calculated for each destination which is the weighted sum of all destinations' individual utilities, with the weighting being provided by the appropriate cell from the probabilistic regionalization matrix. As with the discrete nested logit model, statistically significant weighted regional utility parameter estimates imply that migrants are considering destination characteristics at a regional spatial scale and are therefore evidence that migrants are employing hierarchical destination choice processes.

The hybrid weighted nested logit model is a combination of the competing destinations and weighted nested logit models as it contains the former model's accessibility variables as well as the latter model's weighted regional utility variable.

Results

Both R^2_{adj} and AIC goodness-of-fit statistics for origin-specific calibrations of the competing destinations model show goodness-of-fit improvements over the traditional model for the majority of migration origins. Furthermore, when considering the migration of all adult migrants (in order to maximize sample size) the parameter estimates for the accessibility variable are significant for 99 out of the 100 origins under consideration. Because this accessibility variable is the only operational difference between the traditional and competing destinations models, these results show that it is the inclusion of the accessibility variable representing each destination's likelihood of being cognized within a larger destination cluster, which is causing the improvement in model fit. This suggests that migrants' destination choice behaviour is indeed being influenced by the spatial structure in the destination choice set, and reinforces the hypothesis that migrants are indeed employing hierarchical approaches when selecting destinations.

It was observed that parameter estimates for several of the variables that the competing destinations model has in common with the traditional model were systematically affected by the inclusion of the accessibility variable in the competing destinations model. After detailed consideration it was concluded that this was caused by the moderate correlations that exist between the accessibility variable and several other explanatory variables.

Origin-specific calibrations of the discrete nested logit model provide only marginal improvements in goodness-of-fit over the tradition model, and the discrete regional utility variable's parameter estimate was statistically significant for only 59 of the 100 selected study areas. This is not entirely unexpected because it is intuitively unreasonable to expect any single discrete regionalization to effectively represent the mental maps of all migrants leaving any particular destination. Indeed, it was shown that the goodness-of-fit and parameter estimates from the discrete nested logit model can be very sensitive to the specific discrete regionalization against which the model is calibrated. This intuitive shortcoming of discrete regionalizations and the demonstrated sensitivity of the model were the main motivations for the derivation and application of the weighted nested logit model and the algorithm to generate its associated probabilistic regionalizations. The parameter estimates from origin-specific calibrations of the discrete nested logit model showed very little variation from those of the traditional model.

The goodness-of-fit of origin-specific calibrations of the weighted nested logit model is noticeably better than for both the traditional and the discrete nested logit models, which is not surprising given the high sensitivity of the discrete model's results to the specific discrete regionalization against which it is calibrated. This supports the proposition that the probabilistic approach to regionalization of the weighted nested logit model is more appropriate when modelling aggregate migration behaviour. The weighted regional utility variable is statistically significant for 82 of the 100 selected study areas, suggesting that migrants from many origins are making their migration destination selections in a hierarchical manner. The parameter estimates from origin-specific weighted nested logit calibrations are also very similar to those obtained from the traditional model – there are very few origins for which the difference between any traditional and weighted nested logit parameter estimates are statistically significant.

The predictive performance of the weighted nested logit model is not as good or as widespread as that of the competing destinations model. Detailed comparison of the competing destinations model and the weighted nested logit model suggested that the two models are capturing largely independent hierarchical aspects of migrants' decision-making processes. There is a very low statistical or spatial correlation between the parameter estimates for the accessibility and weighted regional utility variables, and there is also limited correlation between the lists of origins from which each model produces the largest improvements in model fit over the traditional model. For these reasons, the hierarchical variables from the two models were then applied together in the hybrid weighted nested logit model.

It is immediately evident from both goodness-of-fit statistics that the hybrid weighted nested logit model provides by far the best predictive ability of the hierarchical models applied here. For origin-specific calibrations of the hybrid model the accessibility variable's parameter estimates are statistically significant for 99, and the weighted regional utility variable's parameter estimates are statistically significant for 81 of the 100 study areas. The goodness-of-fit of the three best-performing hierarchical models: competing destinations, weighted nested logit and hybrid weighted nested logit, can be compared in figure 11.1 below which plots the AIC statistic for origin-specific calibrations of the three models for the 100 study areas. The superior predictive ability of the hybrid model, which is better than the model fit for almost all of the 100 selected origins, provides considerable evidence in support of a hierarchical theory of migration destination choice, as propounded as the central research question of this thesis.

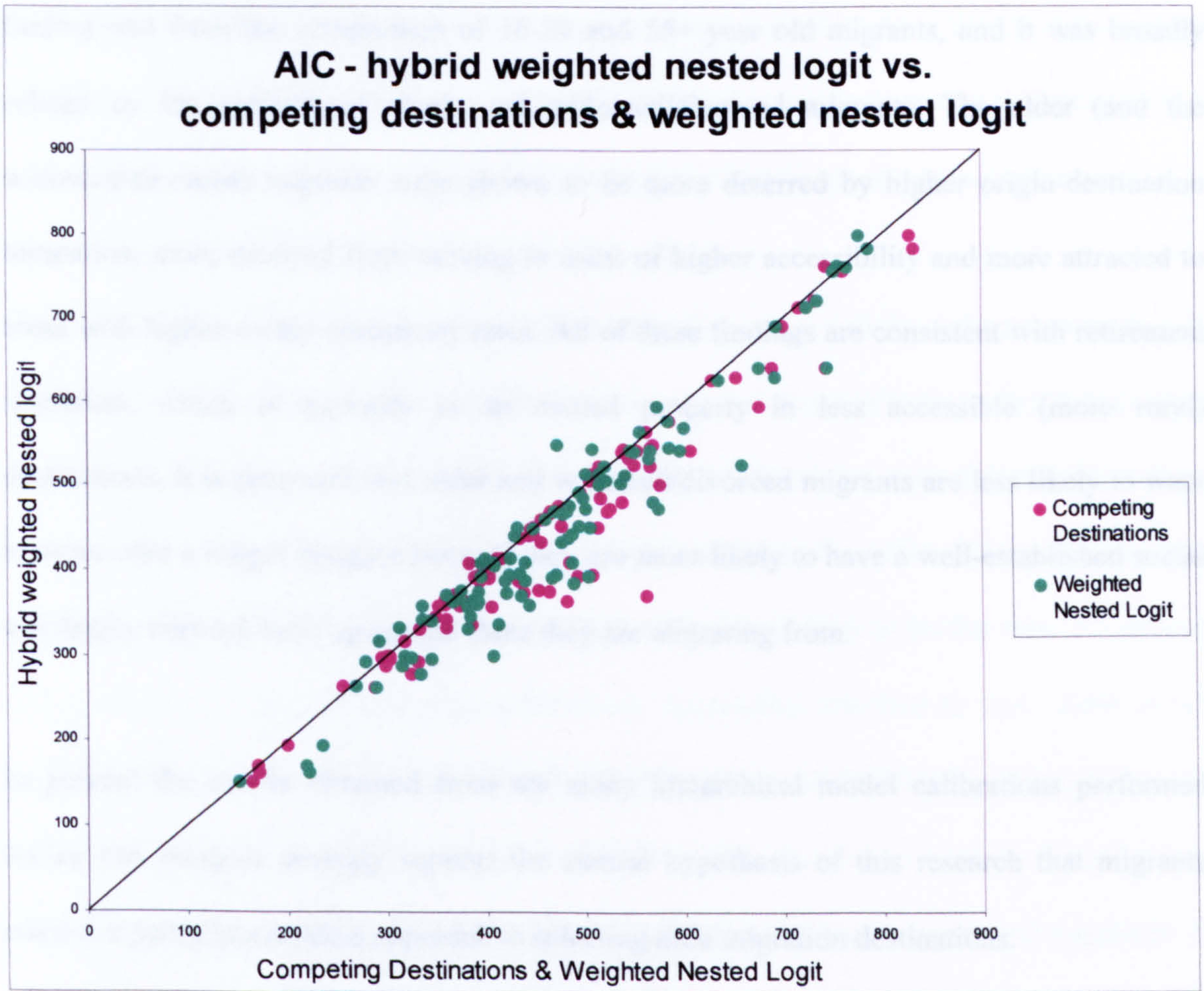


Figure 11.1: AIC goodness-of-fit: hybrid, weighted nested logit and competing destinations.

The parameter estimates from the hybrid model showed skew from those of the traditional model that was consistent with the multicollinearity introduced to the model by the accessibility variable's modest correlations with several of the model's explanatory variables.

Having established the hybrid model as the best performer it was then used to investigate whether there is any significant variation in the migration destination choice behaviour of migrants of different: age, gender, and marital status. The findings from these analyses were hampered by the reduced sample sizes of the various migrant subgroups, which led to higher standard errors on parameter estimates, rendering many statistically insignificant. These analyses showed no notable differences in behaviour between male and female migrants, but did find some differences between age and marital status groups. The most marked of these finding was from the comparison of 16-24 and 55+ year old migrants, and it was broadly echoed by the analysis of single and widowed/divorced migrants. The older (and the widowed/divorced) migrants were shown to be more deterred by higher origin-destination separation, more deterred from moving to areas of higher accessibility and more attracted to areas with higher owner occupancy rates. All of these findings are consistent with retirement migration, which is typically to an owned property in less accessible (more rural) destinations. It is proposed that older and widowed/divorced migrants are less likely to want to move over a longer distance because they are more likely to have a well-established social and family network built up around there they are migrating from.

In general the results obtained from the many hierarchical model calibrations performed during this analysis strongly support the central hypothesis of this research that migrants employ a partly hierarchical approach to selecting their migration destinations.

Limitations of this research

The primary limitation of this research is the nature of the migration data against which the migration models have been calibrated. Availability of suitable aggregate data sets is a perennial problem in social sciences research in general, and in geographical enquiry in particular. This is because of the conflicting requirements to ensure statistical significance and individuals' anonymity through larger sample sizes, whilst also achieving disaggregate relevance through finer spatial resolution.

It is likely that the hierarchical migration destination choice process demonstrated by this research operates at many spatial scales. In this research the use of the district-level Special

Migration Statistics dataset dictates the finest spatial scale at which migration destination choice is examined. Because districts are somewhat large this dataset does not lend itself to the definition of hierarchical destination choice sets with more than 2 spatial scales. It is possible to undertake similar research using inter-ward migration data, though less migrant information is available at that scale. Also it is likely that the number of out-migrants from many wards will be insufficient to produce statistically significant parameter estimates.

Furthermore, the migration system under analysis is a carefully selected set of 100 destination districts that the author believes have a well-defined 'identity', often being based around a single major settlement. This increases the likelihood that each district will form an 'atomic unit' within migrants' cognitive representations of space, rather than a district being split between two cognised destination clusters as is more likely to be the case if a district contains two or more similar size settlements. Restricting analysis to this subset of all districts further exacerbates the problem of three-tier choice set definition.

Another limitation of the current research is its inability to model contracted migration. A large amount of the migration within Great Britain is likely to be job-related. In such situations, spatial choice is constrained by the need to end up within a reasonable commuting distance of a particular location. Contracted migration limits how accurate a two-level nested logit migration model can be, because a large proportion, perhaps the majority, of individual migrations are not influenced by the regional utility component of nested logit models as the destination choice at that broader regional scale is dictated by the location of the job that an individual is moving to.

Suggestions for further research

It would be interesting to compare the results of the analysis presented here with similar analysis performed against observed migration data from the 2001 Census. However, as the primary goal of this research is to demonstrate the hierarchical nature of migrants' decision-making through comparison of the accuracy of hierarchical and traditional migration models, the specific migration dataset against which these models are calibrated is of secondary importance.

Analysis of Census ward-level migration destination choice was not attempted in the current research because it was anticipated that the small flow sizes would lead to statistically insignificant results. However, this is an area that might benefit from further research. In particular, it might be possible to agglomerate ward-level migration data to an intermediate spatial scale, larger than wards but smaller than districts, that would provide sufficient gross out-migration from each area whilst maintaining a meaningful correlation with settlement structure. Such a dataset might be able to support migration analysis using hierarchical destination choice sets with three or more levels.

Such a three-level hierarchical analysis might also be more successful at modelling the migration of contracted migrants. Whilst job location is likely to reduce the significance of a regional utility variable at the broadest spatial scale, regional utility is increasingly likely to become a significant factor at smaller spatial scales because there will be more potential destinations within commuting distance of a contracted migrant's new workplace which will increase the likelihood that they will group destinations and consider some of their characteristics collectively in order to simplify the migration destination choice process. Three-level analysis may, however, experience statistical significance problems, as mentioned above, due to the potentially very low migration flows between smaller areal units.

The operation of hierarchical decision making at a variety of spatial scales could potentially be simulated in the competing destinations model by including a number of accessibility variables calculated using different distance decay exponents. An accessibility statistic generated with a more negative distance decay exponent would still represent the likelihood of a particular destination being cognised in a larger cluster but its values would vary more rapidly across space. This would make it more useful for migration analysis at smaller spatial scales where differentiation between smaller spatial units is necessary.

Ideally, analysis of contracted migration requires observed migration flow data disaggregated by reason for move, such that models can be calibrated independently for the various migrant subgroups categorized on the basis of their migration motivation. There is no current survey which provides such information within Great Britain. Of course, contracted migrants are not the only large group sharing common characteristics that influence their migration behaviour, students away from home at university and members of the armed forced are other valid subgroups as are retirement migrants. Age can be used as a proxy variable for some of these groups, but it is probably not reliable. A migration data set coded by migrant motivation would allow for independent modelling of the differing motivations of such varied migrant subgroups. If such a dataset were ever to become available this could prove to be an interesting and fruitful area of study.

Effective disaggregation of migrants would also allow generated regionalisations to more accurately reflect migrants' cognitive hierarchies. It is likely that the more homogeneous a group of migrants the more similar will be their degree of spatial information and the nature of their cognitive representations of space. Whilst the use of a probabilistic regionalisation to represent migrants' cognised hierarchies simulates the inevitable random variation between individuals' mental maps, these probabilistic regionalisations remain based upon a number of individual discrete regionalisations generated for aggregate migrant groups. They would be even more likely to be representative if their constituent discrete regionalizations were generated independently for more disaggregate migrant groupings.

Another potentially interesting area for research would be the analysis of the migration destination choice behaviour of non-migrants. This may sound nonsensical, but recall that in chapter 2 it was stated that migration analysis comprises two main questions: whether to move, and where to move to. Whilst these questions have traditionally been considered entirely independently it is likely that many non-movers go through some phases of a destination choice process before deciding for whatever reason, perhaps monetary, not to move. Some people who would like to move, but for some reason are not able to, are likely to be making destination choices very frequently, albeit choices that are not acted upon. This is like the Nissan Micra driver who knows exactly which Ferrari he is going to buy when he wins the National Lottery!

Obviously it is not possible to employ a quantitative revealed-preference approach to the analysis of such non-exercised destination choices because there is no aggregate data capturing these 'unrevealed preferences'. Furthermore, this would be a purely academic exercise, as it is only actual migrations that influence future population distribution and that is what the ultimate end-users of these models are, by and large, concerned with.

A potential evolution of the hybrid weighted nested logit model might be to employ different accessibility variables at the two stages of the model's calibration. Because the first results of the first stage calibration are used purely to calculate the regional utility values that are fed into the secondary logit calibration, the first calibration could potentially employ an accessibility variable with lower spatial-autocorrelation, whilst the second stage logit model, which is used to differentiate individual districts (not regions) could employ an accessibility variable that changes more rapidly over space, thus permitting a greater degree of differentiation between potential destinations.

Summary

This research has applied a variety of hierarchical migration models to investigate the destination choice behaviour of those migrants reported by the 1991 Census. The competing destinations and discrete nested logit models each employ different approaches to capture the spatial structure in migrants' destination choice sets and include an explanatory variable to represent this information. The significant parameter estimates observed for both of these explanatory variables signify that migrants are considering spatial structure and aggregate cluster or regional level characteristics when selecting migration destinations.

A novel model, termed the weighted nested logit model, was derived using some probabilistic principles to extend the capabilities of the discrete nested logit model and to make it more intuitively applicable for use when modelling the aggregate behaviour of a diverse group of migrants. Positive results from the application of this model asserted that its probabilistic underpinnings are indeed better suited for this type of analysis than the rigid discrete approach.

Building on the underlying hierarchical assumptions and observed performance of the weighted nested logit and competing destinations models, a hybrid model containing two inherently hierarchical explanatory variables was derived and applied. This model was found to provide the best model-fit to observed Census migration data of all models considered here, further suggesting that the hierarchical assumptions underpinning the models applied in this research are a good approximation to the manner in which actual migrants make their migration destination choice decisions.

The research presented here provides valuable evidence in support of the theory that internal migration destination choice within Great Britain is an inherently hierarchical process. The significantly superior results obtained from hierarchical migration destination choice models that take account of the spatial structure of the destination choice set, compared with the results from traditional migration models, strongly suggests that the hierarchical principles underpinning these more advanced models are a valid and appropriate representation of individual migrants' decision-making behaviour.

Acceptance that there is a hierarchical component to migration destination choice and adoption of some of the techniques for modelling migration destination choice that are presented and applied here, can lead to improvements in the population distribution projections made by many private and public organizations to increase profits (when siting a new retail complex) or even to save lives (when building new fire stations and ambulance dispatch centres).

Bibliography

Allen, G.L. (1981). "A developmental perspective on the effects of 'subdividing' macrospatial experience." Journal of Experimental Psychology: Human Learning and Memory 7: 120-132.

Atkins, D.J. (1995). What? Who? Where? When? and Why? - A survey of migration datasets. Research Report #94/9, North Eastern Regional Research Laboratory, Department of Geography, University of Newcastle upon Tyne.

Atkins, D.J., Champion, A.G., Coombes, M., Dorling, D., Woodward, R. (1996). Migration (report 3). Urban Trends in England: Latest Evaluation from the 1991 Census. T. S. O. B. (Agencies). London, The Stationery Office Books (Agencies).

Atkins, D.J. (1996). Visualising migration: an interactive exploration of 1991 Census data on migration. Geography Department. Newcastle upon Type, University of Newcastle upon Tyne.

Atkins, D.J. and Fotheringham, A.S. (1999). Gender variations in migration destination choice. Migration and Gender in the Developed World. Boyle, P. and Halfacree, K., London, Routledge: 54-72.

Ben-Akiva, M. and Lerman, S.R. (1987). Discrete Choice Analysis - theory and application to travel demand. Cambridge, Mass., The MIT Press.

Boots, B.N. and Kanaroglou, P.S. (1988). "Incorporating the effects of spatial structure in discrete choice models of migration." Journal of Regional Science 28(4): 495-507.

Boyle, P. (1993). Defining migrant profile clusters using data from the 1991 small area statistics and the local base statistics. Research on the 1991 Census, Newcastle, UK.

Boyle, P. (1993). "Modelling the Relationship between Tenure and Migration in England and Wales." Transactions of the Institute of British Geographers **18**(3): 359-376.

Boyle, P. (1995). Changing Faces: contrasting residents and newcomers in remote rural areas. Migration Issues in Rural Areas, Swansea, UK.

Boyle, P. (1995). "Rural in-migration in England and Wales: 1980-81." Journal of Rural Studies **11**(1): 65-78.

Boyle, P., Cooke, T.J., Halfacree, K., and Smith, D. (2001). "A Cross-National Comparison of the Impact of Family Migration on Women's Employment Status." Demography **38**(2): 201-213.

Briggs, R. (1983). Urban cognitive distance. Image and environment: cognitive mapping and spatial behaviour. Downs, R. and Stea, D. Chicago, Aldine: 361-388.

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E. (1998). "Geographically Weighted Regression - Modelling Spatial Non-Stationarity." The Statistician 47(3): 431-443.

Bucklin, R.E.L. and James M. (1991). "A two-state model of purchase incidence and brand choice." Marketing Science **10**: 24-40.

Byrne, R.W. (1979). "Memory for urban geography." Quarterly Journal of Experimental Psychology **31**: 147-154.

Cadwallader, M. (1979). "Problems in cognitive distance: implications for cognitive mapping." Environment and Behaviour **11**: 559-576.

Cameron, L. (2000). "The Residency Decision of Elderly Indonesians: A Nested Logit Analysis." Demography 37(1): 17-27.

Canter, D.V. and Tagg., S.K. (1975). "Distance estimation in cities." Environment and Behavior 7: 59-80.

Carey, H. C. (1858). Principles of Social Science. Philadelphia, Lippincott.

Champion, A.G. (1999). "Migration and British cities in the 1990s." National Institute Economic Review 170: 60-77.

Champion, A.G. (2001). "A Changing Demographic Regime and Evolving Polycentric Urban Regions: Consequences for the Size, Composition and Distribution of City Populations." Urban Studies 38: 657-677.

Champion, A.G. and Atkins, D. J. (2000). Migration between metropolitan and non-metropolitan areas in England and Wales. Migration within England and Wales using the ONS Longitudinal Study. Creeser, R. and Gleave, S., London, The Stationery Office. ONS Series LS no 9: 1-15.

Champion, A.G. and Fielding, A. (1992). Editorial Introduction. Migration Processes and Patterns - volume 1: research progress and prospects. Champion, A.G. and Fielding, A., London, Belhaven Press. 1: 1-14.

Champion, A.G. and Stillwell, J. (1991). Final report: research review and recommendations, IBG Limited Life Working Party on Migration.

Cheng, B. and Titterington, D.M. (1994). "Neural Networks: A Review from a Statistical Perspective." Statistical Science 9(1): 2-30.

Cienki, A.J. (1989). Spatial cognition and the semantics of prepositions in English, Polish and Russian, Munich, Sagner.

Clark, W.A.V. (1993). "Comparing cross-sectional and longitudinal analyses of residential mobility and migration." Environment and Planning A 25: 1291-1302.

CMU (1993). A user guide to the SARs: samples of anonymised records from the 1991 Census of Population of Great Britain. Manchester, Census Microdata Unit, Manchester University.

Congdon, P. (1988). Modelling migration flows between areas: an analysis for London using the Census and OPCS Longitudinal Study. Regional Science Association (Britain) workshop, 08/04/88.

Curtis, A. (1991). An investigation into the hierarchical processing of spatial information by using the competing destination model. Geography. Buffalo, State University of New York: 95.

Curtis, A. and Fotheringham, A.S. (1995). "Large-scale information surfaces: an analysis of city-name recalls in the United States." Geoforum 26(1): 75-87.

Dale, A., Creeser, R., Dodgeon, B., Gleave, S., & Filakti. (1993). "An introduction to the OPCS Longitudinal Study." Environment and Planning A 25: 1387-98.

Desbarats, J. (1983). "Spatial choice and constraints on behaviour." Annals of the Association of American Geographers 73(3): 340-357.

Duke-Williams, O. (1995). Supression in the SMS., Department of Geography, University of Leeds. 1995.

Duke-Williams, O. (1995). Using SMSTAB and SWSTAB, School of Geography, Leeds University.

Evans, G.W. and Pezdek, K. (1980). "Cognitive mapping: knowledge of real-world distance and location information." Journal of Experimental Psychology: Human Learning and Memory 6: 13-24.

Evers, G.H.M. (1990). The residential location and workplace choice: a nested multinomial logit model. Spatial Choices and Processes. M. M. Fischer, P. Nijkamp and Y. Y. Papageorgiou. North-Holland, Elsevier science publishers: 313-329.

Farah, M.J. (1988). The neuropsychology of mental imagery. Spatial Cognition: brain bases and development. J. Stiles-Davis, M. Kritchevsky and U. Bellugi. Hillsdale: New Jersey, Lawrence Erlbaum Associates: 33-56.

Fischer, M.M., Hlavackova-Schindler, K., and Reismann, M. (1999). "A Global Search Procedure for Parameter Estimation in Neural Spatial Interaction Modelling." Journal of Regional Science 78: 119-134.

Fischer, M.M. (2002). "Learning in Neural Spatial Interaction Models: A Statistical Perspective." Journal of Geographical Systems 4: 1-13.

Fischer, M.M. and Gopal, S. (1994). "Artificial neural networks: a new approach to modeling interregional telecommunication flows." Journal of Regional Science 34(4): 503-527.

Fischer, M.M. and Leung, Y. (1998). A Genetic-Algorithms Based Evolutionary Computational Neural Network for Modelling Spatial Interaction Data. European Regional Science Association 38th European Congress, Vienna, Austria.

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

Fotheringham, A.S. (1981). "Spatial Structure and Distance-Decay Parameters." Annals of the Association of American Geographers 71(3): 425-436.

Fotheringham, A.S. (1982). "Distance-Decay Parameters: A Reply." Annals of the Association of American Geographers 72(4): 551-553.

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. (1983). "Further discussion of the Poisson interaction model." Geographical Analysis 15(4): 343-347.

Fotheringham, A.S. (1984). "Spatial flows and spatial patterns." Environment and Planning A 16: 529-543.

Fotheringham, A.S. (1985). "Spatial competition and agglomeration in urban modelling." Environment and Planning A **17**: 213-230.

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

Fotheringham, A.S. (1987). Hierarchical Destination Choice: discussion with evidence from migration in The Netherlands. The Hague, Netherlands Interuniversity Demographic Institute.

Fotheringham, A.S. (1988). "Consumer store choice and choice set definition." Marketing Science 7(3): 299-310.

Fotheringham, A.S. (1991). Migration and Spatial Structure: the development of the competing destinations model. Migration Models: macro and micro approaches. J. Stillwell and P. Congdon. London, Belhaven: 57-72.

Fotheringham, A.S. (1991). Statistical modelling of spatial choice: an overview. Spatial analysis in marketing: theory, methods and applications. Ghosh and Ingene. Greenwich (Connecticut), JAI Press. **Supplement 5**: 95-118.

Fotheringham, A.S. (1997). "Trends in quantitative methods 1: stressing the local." Progress in Human Geography **21**(1): 88-96.

Fotheringham, A.S., Brunsdon, C. and Charlton, M.E. (2000). Quantitative Geography: perspectives on spatial data analysis. London, Sage.

Fotheringham, A.S. and A. Curtis (1992). Encoding Spatial Information: the evidence for hierarchical processing. International conference GIS - From Space to Territory: theories and methods of spatio-temporal reasoning, Pisa, Italy, Springer-Verlag.

Fotheringham, A.S. and Curtis, A. (1995). "Large-scale information surfaces: an analysis of city name recalls in the United States." Geoforum 26: 65-78.

Fotheringham, A.S., Densham, P. J., and Curits, A. (1995). "The zone definition problem in location-allocation modelling." Geographical Analysis 27(1): 60-77.

Fotheringham, A.S. and Dignan, T. (1984). "Further contributions to a general theory of movement." Annals of the Association of American Geographers 74(4): 620-633.

Fotheringham, A.S. and T.C. Pitts (1995). "Directional variation in distance decay." Environment and Planning A 27: 715-729.

Fotheringham, A.S. and Pellegrini, P.A. (1996). "Microdata for migration analysis: a comparison of sources in US, Britain and Canada." Area 28: 347-357.

Franklin, N. (1992). "Spatial representation for describing environments." Geoforum 23(2): 165-174.

Garling, T., Book, A. and Ergezen, N. (1982). "Memory for the spatial layout of the everyday physical environment: differential rates of acquisition of different types of information." Scandinavian Journal of Psychology 23: 23-35.

Golledge, R.G. (1977). Environmental Cues, Cognitive Mapping, and Spatial Behavior. Behavior-Environment Research Methods. D. Burke. Wisconsin, University of Wisconsin, Institute of Environmental Studies: 35-46.

Golledge, R.G. and Zannaras., G. (1973). Cognitive approaches to the analysis of human spatial behavior. Environmental Cognition. W. H. Ittelson. New York, Seminar: 59-94.

Gould, P. and White, R. (1974). Mental Maps. Baltimore, Pengiun.

Guldmann, J-M. (1999). "Competing Destinations and Intervening Opportunities Interaction Models of Inter-City Telecommunications Flows." Papers in Regional Science 78: 179-194.

Guldmann, J-M. (2000). Spatial Interaction Models of International Telecommunications Flows. 40th Congress of European Regional Science Association, Barcelona, Spain.

Halfacree, K. and P. Boyle (1993). "The challenge facing migration research: the case for a bibliographical approach." Progress in Human Geography 17(3): 333-348.

Herskovits, A. (1986). Language and spatial cognition: an interdisciplinary study of prepositions in English. Cambridge, UK, Cambridge University Press.

Hirtle, S.C. and Jonides, J. (1985). "Evidence of hierarchies in cognitive maps." Memory and Cognition 13: 208-217.

Holland, J.H. (1975). Adaptation in Natural and Artificial Systems. Ann Arbor, University of Michigan Press.

Huber, J.J., Payne, W. & Pluto, C. (1982). "Adding a symmetrically dominated alternative: violations of regularity and the similarity hypothesis." Journal of Consumer Research 9: 90-98.

Jedidi, K., Mela, C.F. & Gupta, S. (1999). "Managing advertising and promotion for long-run profitability." Marketing Science 18(1): 1-22.

Jones, K. (1991). "Specifying and estimating multi-level models for geographical research." Transactions of the Institute of British Geographers 16(2): 148-159.

Kitchin, R.M. (1996). "Are there Sex Differences in Geographical Knowledge and Understanding?" The Geographical Journal 162(3): 273-286.

Kitchin, R.M. (1997). "Exploring Spatial Thought." Environment and Behavior 29(1): 123-156.

Kosslyn, S.M. (1984). Mental Imagery Ability. Human abilities: an information-processing approach. Sternberg, R.J., W.H. Freeman and Co., New York.

Kuipers, B., Tecuci, D.G., and Stankiewicz, B.J. (2003). "The Skeleton in the Cognitive Map: a Computational and Empirical Exploration." Environment and Behavior 35(1): 81-106.

Lin, G. and Xie, Y. (1998). "The Log-Linear Modelling of Interstate Migration: Some Additional Considerations." American Sociological Review 63(6): 900-907.

MacEachren, A.M. (1992). "Application of Environmental Learning Theory to Spatial Knowledge Acquisition from Maps." Annals of the Association of American Geographers 82(2): 245-274.

Mackay, D.I. (1969). Geographical mobility and the brain drain: a case study of Aberdeen University graduates, 1860-1960. London, Allen and Unwin.

Maier, G. and Weiss, P. (1990). The discrete choice approach to migration modeling. Migration Models: macro and micro approaches. J. Stillwell and P. Congdon. London, Belhaven Press: 17-33.

Maki, R.H. (1981). "Categorization and distance effects with spatial linear orders." Journal of Experimental Psychology: Human Learning and Memory 7: 15-32.

Mark, D.M. (1992). Counter-intuitive geographical 'facts': clues for spatial reasoning at geographical scales. GIS - From Space to Territory: Theories and methods of spatio-temporal reasoning, Pisa, Italy, Springer-Verlag.

McFadden, D. (1978). Modelling the choice of residential location. Spatial interaction theory and planning models. Karlqvist, Lundqvist, Snickars and Weibull. Amsterdam, North-Holland: 75-96.

McNamara, T.P., Ratcliff, R. and McKoon, G. (1984). "The mental representation of knowledge acquired from maps." Journal of Experimental Psychology: Learning, Memory and Cognition **10**: 723-732.

McNamara, T.P. (1986). "Mental representations of spatial relations." Cognitive Psychology **18**: 87-121.

MIDAS (1997). SMS Gopher Page.

Mitchelson, R.L. and Wheeler, J.O. (1994). "The Flow of Information in a Global Economy: The Role of the American Urban System in 1990." Annals of the Association of American Geographers **84**(1): 87-107.

Newbold, K.B. (1996). "Internal Migration of the Foreign-Born in Canada." International Migration Review 30(3): 728-747.

Newcombe, N. and Liben, L.S. (1982). "Barrier effects in the cognitive maps of children and adults." Journal of Experimental Child Psychology **34**: 46-58.

OPCS (1991). Census User Guide 3: local statistics prospectus. London, OPCS.

OPCS (1991). 1991 Census User Guide 51: Special Migration Statistics: cell numbering layouts, OPCS.

OPCS (1992). 1991 Census User Guide 35: Special Migration Statistics: prospectus. London, OPCS.

Openshaw, S. (1993). Modelling spatial interaction using a neural net. GIS, Spatial Modelling and Policy. Fischer, M.M and Nijkamp, P. Berlin, Springer: 147-164.

Plane D.A. and Mulligan, G.F. (1997). "Measuring spatial focusing in a migration system." Demography 34(2): 251-262.

Plane D.A. and Rogerson, P.A. (1994). The geographical analysis of population with applications to planning and business. New York, John Wiley.

Presson, C.C. and Somerville, S.C. (1985). Beyond egocentrism: a new look at the beginnings of spatial representation. The Development of Children's Spatial Search. H. Wellman. Hillsdale, New Jersey. Erlbaum.

Ravenstein, E.G. (1885). "The Laws of Migration." Journal of the Royal Statistical Society 48(Part 2): 199.

Reggiani, A., Nijkamp, P., and Tsang, W. (1997). Neural Network and Spatial Interaction Analysis of European Commodity Flows. Amsterdam, Holland, Vrije Universiteit Amsterdam - Faculteit der Economosche Wetenschappen en Econometrie.

Sadalla, E. and Magel, S. (1980). "The perception of traversed distance." Environment and Behaviour 12: 65-79.

Sadalla, E. and Staplin, L. (1980). "The perception of traversed distance: intersections." Environment and Behavior 12: 167-182.

Sandberg, E.H., Huttenlocher, J., and Newcombe, N. (1996). "The Development of Hierarchical Representation of Two-Dimensional Space." Child Development 67(3): 721-739.

Sayer, A. (1976). "A critique of urban modelling." Progress in Planning 6: 187-254.

Sayer, A. (1992). Method in Social Science. London, Routledge.

Sherman, R.C., Croxton, J., and Giovanatto, J. (1979). "Investigating Cognitive Representations of Spatial Relationships." Environment and Behavior 11(2): 209-226.

Sholl, M.J. (1987). "Cognitive maps as orienting schemata." Journal of Experimental Psychology: Learning, Memory and Cognition 13: 615-628.

Squire, L.R. (1987). Memory and Brain. New York, Oxford University Press.

Stevens, A. and Coupe, P. (1978). "Distortions in judged spatial relations." Cognitive Psychology 10: 422-437.

Stevens, S.S. (1975). Psychophysics: introduction to its perceptual, neural and social prospects. New York, Wiley.

Talmy, L. (1983). How language structures space. Spatial Orientation: theory, research and application. H. A. Pick, L. New York, Plenum: 225-282.

Thorndyke, P.W. (1981). "Distance estimation from cognitive maps." Cognitive Psychology 13: 526-550.

Thorndyke, P.W. and Hayes-Roth, B. (1982). "Differences in spatial knowledge acquired from maps and navigation." Cognitive Psychology 14: 560-589.

Turton, I., Openshaw, S. and Diplock, G. (1997). A Genetic Programming Approach to Building New Spatial Models Relevant to GIS. Innovations in GIS. Z. Kemp. London, Taylor and Francis. 4: 89-104.

Tversky, B. (1981). "Distortions in memory for maps." Cognitive Psychology 13: 407-433.

West, P.M., Brockett, P.L., and Golden, L.L. (1997). "A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice." Marketing Science 16(4): 370-391.

Wier, T.G. and Phoha, V.V. (1999). "Neural-Network Modeling for Labor-Force Migration" a Competitive-Learning Approach." Computing in Economics and Finance 1999 (Society for Computational Economics) 1043.

Williams, P.A. and A.S. Fotheringham (1984). The calibration of spatial interaction models by maximum likelihood estimation with program SIMODEL. Indiana, US, Indiana University.

Wilson, A.G. (1967). "A statistical theory of spatial distribution models." Transportation Research 1: 253-69.

Wilton, R. (1979). "Knowledge of spatial relations: the specification of the information used in making inferences." Quarterly Journal of Experimental Psychology 31: 133-146.

Wong, D.W.S. and Fotheringham, A.S. (1990). "Urban Systems as Examples of Unbounded Chaos: Exploring the Relationship between Fractal Dimension, Rank-Size, and Rural-to-Urban Migration." Geografiska Annaler 72(2/3): 89-99.

Wood, G. (1983). Cognitive psychology: a skills approach. Belmont, California, Wadsworth.

Wrigley, N., Longley, P., and Dunn, R.. (1988). Some recent developments in the specification, estimation and testing of discrete choice models. Behavioural Modelling in Geography and Planning. Golledge and Timmermans. Beckenham (Kent), Croom-Helm: 96-123.

Young, D.A. (2002). "A New Space-Time Computer Simulation Method for Human Migration." American Anthropologist 104(1): 138-158.

Appendix A: Explanatory variables

This appendix tabulates the values of the explanatory variables for all 100 migration destinations selected for analysis. These variables are described in detail in chapter 4, but in summary:

- The population value is the population aged 16 or over that is usually resident within each selected district on census night 1991. The values of this variable were obtained from the 1991 Census Local Base Statistics.

- House price data is weighted by the composition of the areas' housing stock, and was obtained from Nationwide Building Society.

- The social class variable is defined as the percentage of the workforce of a district that are in professional or managerial positions. The values of this variable were obtained from the 1991 Census Local Base Statistics.

- The tenure variable signifies the percentage of owner occupancy within each selected district. The values of this variable were obtained from the 1991 Census Local Base Statistics.

- Unemployment is the percentage of the workforce without a job on the day of the 1991 census. The values of this variable were obtained from the 1991 Census Local Base Statistics.

- Accessibility is a measure of the spatial competition that a destination district experiences due to its location relative to other districts, and for a particular district is calculated by summing the population/distance for all other district. More isolated districts will have lower accessibility values.

Destination district name	Population	House price	Social class	Tenure	Unemp.	Access.
Aberdeen City	204885	38575	38.93	50.82	4.93	136.86
Barking and Dagenham	143681	62081	18.39	51.76	11.67	780.07
Barnsley	220937	29303	24.87	61.8	12.9	495.31
Bath	78689	90430	43.98	64.74	8.65	392.7
Birmingham	961041	48439	28.01	60.1	14.24	522.14
Blackburn	136612	45222	27.58	69.71	11.49	487.61
Blackpool	146069	53655	28.75	76.1	10.85	378.55

Destination district name	Population	House price	Social class	Tenure	Unemp.	Access.
Bolton	258584	48191	32.66	69.85	10.31	547.74
Bournemouth	151302	61623	37.83	72.62	10.52	335.54
Bradford	457344	46898	32.72	71.22	10.97	455.86
Brighton	143582	76330	42.23	64.59	11.31	397.49
Bristol	376146	54390	35.18	63.75	10.31	379.76
Bury	176760	57032	38.51	76.64	7.37	577.53
Cambridge	91933	77311	48.57	54.26	7.21	434.51
Camden	170444	138675	52.06	33.75	13.19	1002.6
Canterbury	123947	69131	40.55	75.1	8.13	334.86
Cardiff	279055	50619	38.94	69.63	10.96	343.62
Carlisle	100562	39295	32.91	67.12	6.92	244.49
Chelmsford	152418	73746	50.23	78.09	5.5	482.6
Cheltenham	103115	71976	43.85	71.88	7.14	433.12
Chester	115971	70662	45.24	71.57	7.57	443.07
Colchester	142515	68390	39.68	73.75	7.11	376.24
Coventry	294387	45467	29.48	71.05	12.07	506.44
Croydon	313510	82962	43.95	72.77	8.35	754.25
Darlington	98906	25925	33.65	71.12	10.39	326.65
Derby	218802	54669	32.88	68.93	9.8	500.01
Doncaster	288854	37561	26.6	66.92	13.12	435.09
Dover	103216	73083	33.47	70.9	7.81	295.23
Dudley	304615	48777	31.2	68.62	8.97	556.7
Dundee City	165873	34332	30.91	45.45	12.82	180.55
Durham	80669	45703	40.18	61.78	9.2	334.98
Ealing	275257	89903	41.9	63.79	11	837.23
Edinburgh City	418914	46088	44.74	66.15	8.46	223.44
Exeter	98125	74073	34.58	68.02	7.34	251.37
Glasgow City	105202	37942	27.62	37.29	19.08	302.27

Destination district name	Population	House price	Social class	Tenure	Unemp.	Access.
Gloucester	101608	53448	33.17	76.16	7.66	411.69
Greenwich	207650	77621	31.94	47.07	13.29	833.54
Guildford	122378	98803	51.91	73.87	5.02	544.07
Hackney	181248	81814	31.61	26.93	22.49	987.64
Hammersmith and Fulham	148502	106068	46.88	41.95	13.33	997.98
Harrogate	143526	72903	50.04	75.52	4.16	383.08
Harrow	200100	92444	48.44	77.91	7.39	784.66
Ipswich	116956	57125	29.73	65.09	8.22	327.91
Islington	164686	79205	37.75	26.7	17.06	1017.9
Kensington and Chelsea	138394	141240	60.24	39.86	11.87	1055.7
Kings Lynn and West Norfolk (was West Norfolk)	130462	51780	30.87	69.68	7.9	326.55
Kingston upon Hull	254117	41557	21.23	49.38	15.27	332.5
Lambeth	244834	86083	40.12	36.26	17.1	956.1
Lancaster	123856	49910	35.55	76.6	8.06	340.04
Leeds	680722	47910	35.02	61.39	9.18	431.2
Leicester	270493	52497	25.02	57.57	13.76	490.62
Lincoln	81987	45726	28.04	62.34	12.46	385.81
Liverpool	452450	39297	23.67	51.08	21.14	487.48
Luton	171671	60415	28.86	74.13	10.37	543.76
Macclesfield	151590	80270	57.67	76.89	4.96	525.65
Maidstone	136209	78994	44.7	74.49	6.02	439.2
Manchester	404861	34151	27.22	41.26	18.72	587.68
Middlesbrough	140849	34799	28.18	60.9	16.67	336.92
Milton Keynes	176330	56516	38.68	69.18	7.54	479.05
Newbury	136700	79549	48.63	74.09	4.5	447.39
Newcastle upon Tyne	259541	31282	32.94	49.91	15.04	340.8
Newport	133318	39535	33.09	67.56	11.1	369.68
Northampton	180567	54203	33.35	70.5	7.98	463.55

Destination district name	Population	House price	Social class	Tenure	Unemp.	Access.
Norwich	120895	60827	31.43	46.99	10.36	299.15
Nottingham	263522	42648	25.17	51.96	15.38	507.62
Oldham	216531	39968	29.04	67.3	10.74	574.6
Oxford	110103	82194	43.79	57.74	8.51	462.29
Peterborough	153166	51761	32.5	65.15	10.22	395.18
Plymouth	243373	50453	24.89	64	11.04	199.58
Poole	133050	72445	36.45	79.71	7.79	330.69
Portsmouth	174697	64964	30.36	67.1	10.13	389.7
Preston	126082	54719	32.87	67.7	10.52	442.79
Reading	128877	83000	39.68	67.52	7.62	508.55
Rochdale	202164	43759	31.02	64.09	11.63	557.28
Rochester upon Medway	144870	75720	30.46	76.67	9.3	503.03
Rotherham	251637	40090	26.58	61.86	12.01	490.51
Salford	220463	48346	28.89	52.72	13.4	601.75
Scarborough	106221	48327	35.17	71.02	8.21	279.11
Sheffield	501202	42030	32.86	56.78	12.37	480.21
Southampton	196864	59405	30.49	60.94	11.19	393.76
Southwark	218541	83578	32.26	27.17	18.16	965.94
St Albans	126202	97969	59.82	76.04	5.14	619.52
Stafford	117788	48914	44.89	74.44	5.51	479.57
Stirling	78833	50208	45.34	58.07	8.36	214.76
Stockport	284395	58279	45.85	77.8	6.9	568.83
Stoke-on-Trent	244637	34679	19.99	66.35	9.62	483.38
Stratford-on-Avon	105586	79150	50.13	73.18	4.67	468.43
Sunderland	289040	32332	24.88	53.26	14.7	323.06
Swansea	181906	40055	35.74	68.06	11.42	273.1
Thamesdown	170850	63546	34.53	73	7.4	404.79
Tower Hamlets	161064	61921	22.74	23.24	21.8	967.74

Destination district name	Population	House price	Social class	Tenure	Unemp.	Access.
Trafford	212731	59277	42.53	72.86	8.18	588.08
Wakefield	310915	36312	27.74	61.07	10.04	470.21
Walsall	259488	47025	26.07	60.69	11.58	575.37
Warrington	182685	56190	37.86	73.58	7.96	531.53
Warwick	116299	80839	48.47	73.29	6.25	500.04
Wigan	306521	39644	28.71	70.02	10.56	524.46
Wokingham	139189	88539	60.65	85.33	4.27	530.46
Wolverhampton	242190	45242	25.5	57.73	14.29	561.94
York	98745	47184	31.86	64.8	7.71	371.3

Appendix B: Source code for the regionalization algorithm

The source code below can provide a detailed understanding of the algorithm employed to generate discrete regionalizations. This version of the code is written to be compiled and run on a Microsoft Windows machine (there is a dependency on some windows scripting commands). Clarity of purpose rather than performance optimization was the primary design goal for this software.

```
/* This program reger.c allocates the 459 local authority districts of Great Britain
to generate regionalisations which can form the input to a Nested Logit model of
migration destination choice.

Separate regionalisations are generated for each selected migration origin. The number
of regions in the regionalisation is determined by the accessibility of each
individual origin district.

A predetermined number (NumRegionalisations) of regionalisations are written to file
for each origin (as a list of seed district numbers followed by the standardised
inter-region information variance of that regionalisation), as long as this number of
valid regionalisations is generated within MaxIters iterations of the seed district
selection routine. In this case, the regionalisation generator moves onto the next
origin district after NumIters iterations.

The information a migrant from origin I has about destination J is calculated as:
  (pop_of_J * PopExp) * (distance_I_to_J * DistExp)
PopExp and DistExp are global variables, (they may be command line settable in
future).

Generated regions are constrained to contain at least:
  a) a minimum number of districts, and
  b) a minimum number of potential destinations (i.e. districts in selected migration
system).

Ver 3.0       03/12/99             David Atkins
*/

/* Include library files here */
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>
#include <time.h>

/* Global configuration #define's and variables in here     */
#define NumRegionalisations 500000   /* Num of valid regionalisation to generate per
                                     origin  */
#define PopExp 1.0                   /* Population exponent in 'Info. function' */
#define MaximumIterations 1000000000 /* Num of seeds randomly selected per origin */
#define DistExp -1.0                 /* Distance exponent in 'Info. function' */
#define MinNumAny 10                 /* Minimum number of any districts per region */
#define MinNumSelected 7             /* Min number of selected districts per region */
#define StartOrigin 61               /* Nested Logit order number of starting origin
*/
#define NumberOrigins 1              /* Number origins to process */
#define MatrixPercentage 10          /* % of top regionalizations for Weighted matrix
*/

/* Other global variables */
long district_data[460][3];          /* Population, easting, northing */
double calcSeparation();             /* Function to calculate district separation */

int main(argc, argv)
int argc;
char argv[];
{
  /* First declare all variables     */
  long temp_time, start_time, finish_time;
  time_t dummy_time;
  double district_info[460], region_info[12];
```

```
  double district_accessibility[460], calcAccessibility(), min_accessibility,
max_accessibility;
  double incremental_accessibility;
  double separation, calcSeparation(), closest_distance;
  double sum_of_squares, mean, sum, random();
  double *variances;
  short district_num_regions[460];
  short district_nl_index[460];
  short (*regionalisations)[13];
  short my_random[12], district_region[460];
  unsigned random_seed;
  short (*regionalisation_region_any_sizes)[12],
(*regionalisation_region_selected_sizes)[12];
  short current_seeds[12], origin_num, closest_seed_num, temp_flag,
number_origins_processed;
  long loop1, loop2, loop3;
  long iterations_done, num_regionalisations_produced;
  long inclusion_matrix[460][460];

  short district_order_info_ranking[460];     /* In OPCS district order (contains index
in
                                              info rank) */
  short info_order_district_index[460];       /* In information order (contains
district
                                              number) */

  char district_codes[460][5], in_line[30], out_filename[50];
  FILE *fopen(),*file;
  void initialiseNLAreas();

  /* Dynamically allocate memory for variances[] and regionalisations[][13] */
  regionalisations = (short (*)[13]) malloc(NumRegionalisations * 13 * sizeof(short)
);
  variances = (double *) malloc(NumRegionalisations * sizeof(double) );
  regionalisation_region_any_sizes = (short (*)[12]) malloc(NumRegionalisations * 12 *
sizeof(short) );
  regionalisation_region_selected_sizes = (short (*)[12]) malloc(NumRegionalisations *
12 * sizeof(short) );

  /* Read in data to global dist_data[][]: code, population, easting, northing    */
  if ((file=fopen("distdata.csv","r"))==NULL)
    return(printf("\n\nFile distdata.csv not found ... terminating\n\n"));
  for(loop1=1;loop1<460;loop1++)
  {
    fgets(in_line,30,file);
    sscanf(in_line,"%4c%*c%i%*c%i%*c%i%*c\n",district_codes[loop1],

        &district_data[loop1][0],&district_data[loop1][1],&district_data[loop1][2]);
  }
  fclose(file);

  printf("Read in district data from distdata.csv\n");

  /* Set district_nl_index (which areas to make regionalisations for)       */
  initialiseNLAreas(district_nl_index);

  /* Calculate the number of regions to be in each NL district's regionalisations. */
  min_accessibility = max_accessibility = calcAccessibility(2);
  for(loop1=1;loop1<460;loop1++)
  {
    if (district_nl_index[loop1]>0)
    {
      district_accessibility[loop1]=calcAccessibility(loop1);
      if (district_accessibility[loop1]<min_accessibility)
        min_accessibility=district_accessibility[loop1];
      else if (district_accessibility[loop1]>max_accessibility)
        max_accessibility=district_accessibility[loop1];
    }
  }
  incremental_accessibility = (max_accessibility-min_accessibility)/4.0;
  for(loop1=1;loop1<460;loop1++)
  {
    if (district_nl_index[loop1]>0)
    {
      district_num_regions[loop1]=6 +
          (int)( (district_accessibility[loop1]-
min_accessibility)/incremental_accessibility );
      printf("district_num_regions[loop1]=%i\n",district_num_regions[loop1]);
    }
  }
```

```
  /* Seed the random number generator */
  temp_time = (long) time(&dummy_time);
  random_seed=(unsigned)(temp_time%1000);
  srand(random_seed);

  printf("Finished all initialisation, starting regionalisation generation...\n\n");

  /*********************************************
  * MAIN ORIGIN PROGRESSION LOOP STARTS HERE   *
  *********************************************/
  loop1=1;
  while (district_nl_index[loop1] != StartOrigin)
    loop1++;
  if (loop1<460)
  {
    origin_num = (short)loop1;
    printf("Setting origin_num = %d\n",origin_num);
  }
  else
    return (printf("Invalid origin specified in '#define Origins' statement...
terminating\n\n"));

  number_origins_processed = 0;

  while(origin_num<460)
  {
    if (district_nl_index[origin_num]>0)
    {
      printf("Producing %i %i-region regionalisations for district
%s\n\n",NumRegionalisations,
                                                    district_num_regions[origin_num],
district_codes[origin_num]);
      start_time = (long) time(&dummy_time);

      /* Initialise data structures */
      for(loop1=0;loop1<NumRegionalisations;loop1++)
      {
        variances[loop1]=0.0;
        for(loop2=0;loop2<13;loop2++)
        {
          regionalisations[loop1][loop2]=0;
          regionalisation_region_any_sizes[loop1][loop2]=0;
          regionalisation_region_selected_sizes[loop1][loop2]=0;
        }
      }
      for(loop1=1;loop1<460;loop1++)
        for(loop2=1;loop2<460;loop2++)
          inclusion_matrix[loop1][loop2]=0;

      /* Calculate info scores for each destination district        */
      for(loop1=1;loop1<460;loop1++)
      {
        if (loop1 != origin_num)
        {
          separation=calcSeparation(loop1,origin_num);

district_info[loop1]=pow((double)district_data[loop1][0],PopExp)*(pow(separation,DistE
xp));
        }
        else
        {
          district_info[loop1]=0;
        }
      }

      /* Rank districts according to their info. score      */
      for(loop1=1;loop1<460;loop1++)
        district_order_info_ranking[loop1]=info_order_district_index[loop1]=0;
      /* Cycle through incrementing rank every time a higher score is found */
      for(loop1=1;loop1<460;loop1++)
        for(loop2=1;loop2<460;loop2++)
          if(district_info[loop2]>district_info[loop1])
district_order_info_ranking[loop1]++;
      /* Also order districts by increasing information to speedup lookup  */
      for(loop1=1;loop1<460;loop1++)
        info_order_district_index[district_order_info_ranking[loop1]]=(short)loop1;

      num_regionalisations_produced=0;
      iterations_done=0;
      /*********************************************
      * 'RANDOM ITERATIONS' LOOP STARTS HERE *
```

```
*******************************************/
      while( (iterations_done<MaximumIterations) &&
(num_regionalisations_produced<NumRegionalisations) )
      {
        if (iterations_done%1000 == 0)
          printf("Completed %d iterations in %d seconds, %d valid regionalisations so
far.\n",
                iterations_done, time(&dummy_time)-start_time,
num_regionalisations_produced);
        /* Semi-randomly select appropriate number of seed district:-
        Based on number of regions required select following number of seed districts
from each quartile Q1-Q4 of the district list when ranked by regional information
variance:
        Regs Q1   Q2   Q3   Q4
        6    2    2    1    1
        7    3    2    1    1
        8    3    2    2    1
        9    4    2    2    1
        10   4    3    2    1
        11   4    3    2    2
        12   4    4    2    2
            +0  +115 +230 +345
        Treat Q4 as having 114 districts, Q1-Q3 having 115 each. */

        /* Initialise random numbers. */
        for (loop1=0;loop1<12;loop1++)
          my_random[loop1]=0;

        /* Generate required number of random numbers. */
        my_random[0]=(int)(random()*113) + 1;
        for (loop1=1;loop1<district_num_regions[origin_num];loop1++)
          my_random[loop1]=(int)(random()*114) + 1;

        /* Use random numbers to generate seed districts. */
        my_random[0]+=345;    /* This random number always represents a Q4 seed */
        if ( (district_num_regions[origin_num] == 6) ||
(district_num_regions[origin_num] == 7) )
        {
          my_random[1]+=230;
          my_random[2]+=115;
          my_random[3]+=115;
        }
        else if ( (district_num_regions[origin_num] == 8) ||
(district_num_regions[origin_num] == 9) )
        {
          my_random[1]+=230;
          my_random[2]+=230;
          my_random[3]+=115;
          my_random[4]+=115;
        }
        else if (district_num_regions[origin_num] == 10)
        {
          my_random[1]+=230;
          my_random[2]+=230;
          my_random[3]+=115;
          my_random[4]+=115;
          my_random[5]+=115;
        }
        else if (district_num_regions[origin_num] == 11)
        {
          my_random[1]+=345;
          my_random[2]+=230;
          my_random[3]+=230;
          my_random[4]+=115;
          my_random[5]+=115;
          my_random[6]+=115;
        }
        else if (district_num_regions[origin_num] == 12)
        {
          my_random[1]+=345;
          my_random[2]+=230;
          my_random[3]+=230;
          my_random[4]+=115;
          my_random[5]+=115;
          my_random[6]+=115;
          my_random[7]+=115;
        }
        else
        {
          printf("Invalid district_num_regions[%i]=%i\n\n",loop1,
district_num_regions[loop1]);
```

```
        }
        for(loop1=0;loop1<district_num_regions[origin_num];loop1++)
        {
          current_seeds[loop1]=info_order_district_index[my_random[loop1]];
        }

        /* Allocate all districts to ONE region and update region frequencies*/
        for(loop1=0;loop1<district_num_regions[origin_num];loop1++)
        {
          region_info[loop1]=0;
          regionalisation_region_any_sizes[num_regionalisations_produced][loop1]=0;

regionalisation_region_selected_sizes[num_regionalisations_produced][loop1]=0;
        }
        for(loop1=1;loop1<460;loop1++)
            {
          closest_seed_num=-1;
          closest_distance=2000000.0;
          for(loop2=0;loop2<district_num_regions[origin_num];loop2++)
          {

if((separation=calcSeparation(loop1,current_seeds[loop2]))<closest_distance)
                        {
              closest_seed_num=(short)loop2;
              closest_distance=separation;
                        }
          }
          if (closest_seed_num >= 0)
          {
            district_region[loop1]=closest_seed_num;
            region_info[closest_seed_num]+=district_info[loop1];
            regionalisation_region_any_sizes[num_regionalisations_produced]
                                                            [closest_seed_num]+
                                                    +;
            if (district_nl_index[loop1]>0)
              regionalisation_region_selected_sizes[num_regionalisations_produced]
                                                            [closest_seed_num]+
                                                    +;
          }
          else
          {
            return(printf("District %i not allocated to a region... terminating",
loop1));
          }
              }

        /* Check all regions contain at least MinNumSelected and MinNumAny districts
*/
        temp_flag=0;
        for(loop1=0;loop1<district_num_regions[origin_num];loop1++)
          if( (regionalisation_region_any_sizes[num_regionalisations_produced][loop1]
< MinNumAny) ||

(regionalisation_region_selected_sizes[num_regionalisations_produced][loop1] <
MinNumSelected) )
              temp_flag=1;

        if (temp_flag==0)
        {
          /* Calculate variance of regional info. totals    */
          sum_of_squares=mean=sum=0;
          for(loop1=0;loop1<district_num_regions[origin_num];loop1++)
            sum+=(double)region_info[loop1];
          mean=sum/((double)district_num_regions[origin_num]);
          for(loop1=0;loop1<district_num_regions[origin_num];loop1++)
            sum_of_squares+=pow((double)region_info[loop1],2.0);
          variances[num_regionalisations_produced]=(
sum_of_squares/((double)district_num_regions[origin_num]) )

- pow(mean,2.0);

          for(loop1=0;loop1<district_num_regions[origin_num];loop1++)

regionalisations[num_regionalisations_produced][loop1]=current_seeds[loop1];

          num_regionalisations_produced++;
        }
        iterations_done++;

      } /* End of RANDOM ITERATIONS LOOP     */
```

```
      /* Sort all_scores, put lowest variance sets into top_scores */
      for(loop1=0;loop1<num_regionalisations_produced;loop1++)
        regionalisations[loop1][12]=0;
      for(loop1=0;loop1<num_regionalisations_produced;loop1++)
        for(loop2=0;loop2<num_regionalisations_produced;loop2++)
          if(variances[loop2]<variances[loop1]) regionalisations[loop1][12]++;

      time(&finish_time);

      /* Write OrigCode, top seed areas numbers & abs_variances to file    */
     sprintf(out_filename,"%s.regionalisations",district_codes[origin_num]);
      if((file=fopen(out_filename,"w"))==NULL)
        return(printf("\n\nCouldn't open file %s for regionalisation output...
terminating.\n\n", out_filename));
      for(loop1=0;loop1<num_regionalisations_produced;loop1++)
      {
        for(loop2=0;loop2<num_regionalisations_produced;loop2++)
        {
          if (regionalisations[loop2][12]==loop1)
          {
            for(loop3=0;loop3<district_num_regions[origin_num];loop3++)
            {

printf("%d{%d,%d},",regionalisations[loop2][loop3],regionalisation_region_any_sizes[lo
op2][loop3],

regionalisation_region_selected_sizes[loop2][loop3]);
              fprintf(file,"%d,",regionalisations[loop2][loop3]);
            }
            printf("%.0f\n",variances[loop2]);
            fprintf(file,"%.0f\n",variances[loop2]);
          }
        }
      }
      fclose(file);

      /* Generate inclusion_matrix */
      for(loop1=0;loop1<num_regionalisations_produced;loop1++)
        if (regionalisations[loop1][12]< ( (int)(
(double)num_regionalisations_produced * (MatrixPercentage/100.0) ) ) )
        {
          for(loop2=1;loop2<460;loop2++)
          {
            closest_seed_num=-1;
            closest_distance=2000000.0;
            for(loop3=0;loop3<district_num_regions[origin_num];loop3++)

if((separation=calcSeparation(loop2,regionalisations[loop1][loop3]))<closest_distance)
                          {
                closest_seed_num=(short)loop3;
                closest_distance=separation;
                          }
            district_region[loop2]=closest_seed_num;
            }

          for(loop2=1;loop2<460;loop2++)
            for(loop3=1;loop3<460;loop3++)
              if (district_region[loop2] == district_region[loop3])
                inclusion_matrix[loop2][loop3]++;
        }

      /* Output inclusion matrix to file */
      sprintf(out_filename,"%s.matrix",district_codes[origin_num]);
      if((file=fopen(out_filename,"w"))==NULL)
        return(printf("\n\nCouldn't open file %s for matrix output...
terminating.\n\n", out_filename));
      for(loop1=1;loop1<460;loop1++)
      {
        for(loop2=1;loop2<459;loop2++)
          fprintf(file,"%d,",inclusion_matrix[loop1][loop2]);
        fprintf(file,"%d\n",inclusion_matrix[loop1][459]);
      }
      fclose(file);

      printf("\n%i regionalisations produced for origin %4s in %ld seconds from %i
iterations.\n",

num_regionalisations_produced,district_codes[origin_num],(finish_time-
start_time),iterations_done);
```

```
      printf("Inclusion matrix generated from top %d%% (%d)
regionalisations.\n",MatrixPercentage,

(int)((double)num_regionalisations_produced*(MatrixPercentage/100.0)) );
    }
    else // district_nl_index <= 0
    {
      printf("Skipping district %s as not a selected NL origin district\n\n",
district_codes[origin_num]);
    }

    number_origins_processed++;

    if ( (number_origins_processed < NumberOrigins) && (StartOrigin +
number_origins_processed <= 100) )
    {
      loop1=1;
      while (district_nl_index[loop1]!=(StartOrigin+number_origins_processed))
        loop1++;
      if (loop1<460)
      {
        origin_num = (short)loop1;
        printf("Setting origin_num = %d\n",origin_num);
      }
    }
    else
      origin_num = 460;
  } /* End of MAIN ORIGIN PROGRESSION LOOP    */
  printf("Completed generating regionalisations for selected migration origin
districts.\n\n");
  return 1;
}

double calcSeparation(dist1,dist2)
short dist1,dist2;
{
  double dist1east,dist1north,dist2east,dist2north;

  dist1east=(double)district_data[dist1][1];
  dist1north=(double)district_data[dist1][2];
  dist2east=(double)district_data[dist2][1];
  dist2north=(double)district_data[dist2][2];
  return(sqrt(pow(dist1east-dist2east,2.0)+pow(dist1north-dist2north,2.0)));
}

double calcAccessibility(districtNumber)
int districtNumber;
{
  int loop;
  double ret_val=0.0;
  for(loop=1;loop<460;loop++)
    if (districtNumber != loop)
      ret_val += pow((double)(district_data[loop][0]),PopExp) *
pow(calcSeparation(districtNumber, loop),DistExp);
  return ret_val;
}

double random()
{
        return (rand()/(double)RAND_MAX);
}

void initialiseNLAreas(nl_areas)
short *nl_areas;
{
  int loop;
  for(loop=0;loop<460;loop++)
    nl_areas[loop]=0;

  nl_areas[418]=1;
  nl_areas[15]=2;
  nl_areas[49]=3;
  nl_areas[70]=4;
  nl_areas[58]=5;
  nl_areas[241]=6;
  nl_areas[242]=7;
  nl_areas[34]=8;
  nl_areas[141]=9;
  nl_areas[65]=10;
  nl_areas[157]=11;
  nl_areas[71]=12;
```

```
nl_areas[35]=13;
nl_areas[91]=14;
nl_areas[2]=15;
nl_areas[228]=16;
nl_areas[398]=17;
nl_areas[118]=18;
nl_areas[168]=19;
nl_areas[178]=20;
nl_areas[97]=21;
nl_areas[169]=22;
nl_areas[59]=23;
nl_areas[20]=24;
nl_areas[150]=25;
nl_areas[125]=26;
nl_areas[50]=27;
nl_areas[230]=28;
nl_areas[60]=29;
nl_areas[455]=30;
nl_areas[152]=31;
nl_areas[21]=32;
nl_areas[432]=33;
nl_areas[132]=34;
nl_areas[445]=35;
nl_areas[181]=36;
nl_areas[23]=37;
nl_areas[341]=38;
nl_areas[3]=39;
nl_areas[4]=40;
nl_areas[293]=41;
nl_areas[24]=42;
nl_areas[334]=43;
nl_areas[6]=44;
nl_areas[7]=45;
nl_areas[274]=46;
nl_areas[223]=47;
nl_areas[8]=48;
nl_areas[247]=49;
nl_areas[68]=50;
nl_areas[259]=51;
nl_areas[266]=52;
nl_areas[45]=53;
nl_areas[76]=54;
nl_areas[102]=55;
nl_areas[233]=56;
nl_areas[36]=57;
nl_areas[107]=58;
nl_areas[88]=59;
nl_areas[81]=60;
nl_areas[54]=61;
nl_areas[382]=62;
nl_areas[282]=63;
nl_areas[276]=64;
nl_areas[305]=65;
nl_areas[37]=66;
nl_areas[308]=67;
nl_areas[95]=68;
nl_areas[135]=69;
nl_areas[145]=70;
nl_areas[192]=71;
nl_areas[249]=72;
nl_areas[82]=73;
nl_areas[38]=74;
nl_areas[234]=75;
nl_areas[51]=76;
nl_areas[39]=77;
nl_areas[296]=78;
nl_areas[52]=79;
nl_areas[194]=80;
nl_areas[11]=81;
nl_areas[211]=82;
nl_areas[328]=83;
nl_areas[410]=84;
nl_areas[40]=85;
nl_areas[330]=86;
nl_areas[353]=87;
nl_areas[57]=88;
nl_areas[403]=89;
nl_areas[365]=90;
nl_areas[12]=91;
nl_areas[42]=92;
nl_areas[69]=93;
```

```
  nl_areas[63]=94;
  nl_areas[104]=95;
  nl_areas[354]=96;
  nl_areas[43]=97;
  nl_areas[85]=98;
  nl_areas[64]=99;
  nl_areas[298]=100;
}
```

Appendix C: Model Automation 'Scripts'

A number of programs were written in C to automate the process of running many set of origin-specific models for a variety of migrant sub-groups. The source code of these automation programs is included below in order that the results presented in this research can be reproduced and also to enable the models employed here may form the basis for future research.

All of the model automation scripts in this appendix make use of the SIModel package for calibration of Poisson regression models. The Fortran source code for the SIModel program can be found in Williams and Fotheringham (1984).

Traditional model

```
/* This is a C program to automatically generate and run global and origin specific
traditional style models for
DJA's PhD research. Traditional models are calibrated globally and origin-specifically
for:
16-24 year old migrants, 25-54 year old migrants and migrants aged 55+.

Must be run from the directory containing simodel.exe

David Atkins    29-01-00              ver.1.0        */

#include <stdio.h>

#define START_AREA 1
#define STOP_AREA 100
#define STEP_BY 1

int loop_num;

main()
{
  int interaction,district,destination,sub_group;
  char systemcall[200],inString[100],filename[200],group_string[5];
  char home[24]="/home/dja/Analysis/";
  FILE *in_flows,*out_flows;

  /* ALL GLOBAL CALIBRATIONS */

  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Population/population100.csv >
%sModels/Traditional/trad_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/SocialClass/socialclass100.csv
>> %sModels/Traditional/trad_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/HousePrices/houseprices100.csv
>> %sModels/Traditional/trad_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Tenure/tenure100.csv >>
%sModels/Traditional/trad_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Unemployment/unemployment100.csv
>> %sModels/Traditional/trad_tail.txt",home,home);
  system(systemcall);

  sprintf(systemcall,"rm -r %sModels/Traditional/LatestRun",home);
```

```
 system(systemcall);
 sprintf(systemcall,"mkdir %sModels/Traditional/LatestRun",home);
 system(systemcall);

 /* AGE-GROUP PROGRESSION LOOP */
 for(sub_group=0;sub_group<12;sub_group++)
 {
   if (sub_group == 0)
   {
     strcpy(group_string,"all_1624");
   }
   else if (sub_group == 1)
   {
     strcpy(group_string,"all_2554");
   }
   else if (sub_group == 2)
   {
     strcpy(group_string,"all_55+");
   }
   else if (sub_group == 3)
   {
     strcpy(group_string,"all_16+");
   }
   else if (sub_group == 4)
   {
     strcpy(group_string,"male_1624");
   }
   else if (sub_group == 5)
   {
     strcpy(group_string,"male_2554");
   }
   else if (sub_group == 6)
   {
     strcpy(group_string,"male_55+");
   }
   else if (sub_group == 7)
   {
     strcpy(group_string,"male_16+");
   }
   else if (sub_group == 8)
   {
     strcpy(group_string,"female_1624");
   }
   else if (sub_group == 9)
   {
     strcpy(group_string,"female_2554");
   }
   else if (sub_group == 10)
   {
     strcpy(group_string,"female_55+");
   }
   else
   {
     strcpy(group_string,"female_16+");
   }

    printf("\nDoing global traditional runs\n");
    /* Competing destinations */
    sprintf(systemcall,"mkdir %sModels/Traditional/LatestRun/%s",home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/Traditional/trad_head.txt >
%sModels/Traditional/LatestRun/%s/global.run",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable3/%s.txt >>
%sModels/Traditional/LatestRun/%s/global.run",

home,group_string,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/Traditional/trad_tail.txt >>
%sModels/Traditional/LatestRun/%s/global.run",home,home,group_string);
    system(systemcall);

    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/Traditional/LatestRun/%s/global.run
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/Traditional/LatestRun/%s/global.res",home,home,group_string);
```

```
    system(systemcall);

    printf("\nFinished global Traditional runs\n");

    /* Perform ORIGIN-SPECIFIC calibrations */
    /* Generate and calibrate run files for OS models */

    printf("\nGenerating OS Traditional data for %s year old
migrants\n",group_string);
    sprintf(filename,"%sData/FlowData/SMSTable3/%s.txt",home,group_string);
    if ((in_flows=fopen(filename,"r"))==NULL)
    {
      return(printf("Error opening input flow file\n"));
    }
    sprintf(filename,"%sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    if ((out_flows=fopen(filename,"w"))==NULL)
    {
      return(printf("Error opening output flow file\n"));
    }
    while(fgets(inString,100,in_flows))
    {
      sscanf(inString,"%10i%10i%*5i%5i",&interaction,&district,&destination);
      fprintf(out_flows,"%10i%10i    1%5i\n",interaction,district,destination);
    }
    fclose(in_flows);
    fclose(out_flows);

    /* MAIN ORIGIN PROGRESSION LOOP */
    for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
    {
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable3/%s.os.txt >
%sData/FlowData/SMSTable3/os.tmp",
            (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/Traditional/trad_head.os.txt >
%sModels/Traditional/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable3/os.tmp >>
%sModels/Traditional/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/Traditional/trad_tail.txt >>
%sModels/Traditional/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);

      printf("\nCompeting destinations run for %s year old migrants from origin
%i\n",group_string,loop_num);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/Traditional/LatestRun/%s/%i.run
%sModels/SIModel/simodel2_0.dat",
              home,group_string,loop_num,home);
      system(systemcall);
      sprintf(systemcall,"%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/Traditional/LatestRun/%s/%i.res",
              home,home,group_string,loop_num);
      system(systemcall);
    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable3/os.tmp",home);
    system(systemcall);
    sprintf(systemcall,"rm %sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    system(systemcall);
        printf("Completed all calibrations for group: %s/n/n",group_string);
  } /* END OF sub_group PROGRESSION LOOP */

  /* MARITAL STATUS PROGRESSION LOOP */
  for(sub_group=0;sub_group<9;sub_group++)
  {
    if (sub_group == 0)
    {
      strcpy(group_string,"all_single");
    }
    else if (sub_group == 1)
    {
      strcpy(group_string,"all_married");
    }
```

```
else if (sub_group == 2)
{
  strcpy(group_string,"all_wid_div");
}
else if (sub_group == 3)
{
  strcpy(group_string,"male_single");
}
else if (sub_group == 4)
{
  strcpy(group_string,"male_married");
}
else if (sub_group == 5)
{
  strcpy(group_string,"male_wid_div");
}
else if (sub_group == 6)
{
  strcpy(group_string,"female_single");
}
else if (sub_group == 7)
{
  strcpy(group_string,"female_married");
}
else
{
  strcpy(group_string,"female_wid_div");
}

printf("\nDoing global traditional runs\n");
/* Competing destinations */
sprintf(systemcall,"mkdir %sModels/Traditional/LatestRun/%s",home,group_string);
system(systemcall);
sprintf(systemcall,"cat %sModels/Traditional/trad_head.txt >
%sModels/Traditional/LatestRun/%s/global.run",home,home,group_string);
system(systemcall);
sprintf(systemcall,"cat %sData/FlowData/SMSTable4/%s.txt >>
%sModels/Traditional/LatestRun/%s/global.run",

home,group_string,home,group_string);
system(systemcall);
sprintf(systemcall,"cat %sModels/Traditional/trad_tail.txt >>
%sModels/Traditional/LatestRun/%s/global.run",home,home,group_string);
system(systemcall);

sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
system(systemcall);
sprintf(systemcall,"cp %sModels/Traditional/LatestRun/%s/global.run
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
system(systemcall);
sprintf(systemcall, "%sModels/SIModel/simodel",home);
system(systemcall);
sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/Traditional/LatestRun/%s/global.res",home,home,group_string);
system(systemcall);

printf("\nFinished global Traditional runs\n");

/* Perform ORIGIN-SPECIFIC calibrations */
/* Generate and calibrate run files for OS models */

printf("\nGenerating OS Traditional data for %s year old
migrants\n",group_string);
sprintf(filename,"%sData/FlowData/SMSTable4/%s.txt",home,group_string);
if ((in_flows=fopen(filename,"r"))==NULL)
{
  return(printf("Error opening input flow file\n"));
}
sprintf(filename,"%sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
if ((out_flows=fopen(filename,"w"))==NULL)
{
  return(printf("Error opening output flow file\n"));
}
while(fgets(inString,100,in_flows))
{
  sscanf(inString,"%10i%10i%*5i%5i",&interaction,&district,&destination);
  fprintf(out_flows,"%10i%10i    1%5i\n",interaction,district,destination);
}
fclose(in_flows);
fclose(out_flows);
```

```
    /* MAIN ORIGIN PROGRESSION LOOP */
    for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
    {
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable4/%s.os.txt >
%sData/FlowData/SMSTable4/os.tmp",
            (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/Traditional/trad_head.os.txt >
%sModels/Traditional/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable4/os.tmp >>
%sModels/Traditional/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/Traditional/trad_tail.txt >>
%sModels/Traditional/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);

      printf("\nCompeting destinations run for %s year old migrants from origin
%i\n",group_string,loop_num);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/Traditional/LatestRun/%s/%i.run
%sModels/SIModel/simodel2_0.dat",
              home,group_string,loop_num,home);
      system(systemcall);
      sprintf(systemcall,"%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/Traditional/LatestRun/%s/%i.res",
              home,home,group_string,loop_num);
      system(systemcall);
    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable4/os.tmp",home);
    system(systemcall);
    sprintf(systemcall,"rm %sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
    system(systemcall);

      printf("Completed all calibrations for group: %s/n/n",group_string);
  } /* END OF sub_group PROGRESSION LOOP */

  printf("Finished\n\n");

  return;
}
```

Competing destinations model

```
/* This is a C program to automatically generate and run global and origin specific
CompetingDestinations style models for
DJA's PhD research. CompetingDestinations models are calibrated globally and origin-
specifically for:
16-24 year old migrants, 25-54 year old migrants and migrants aged 55+.

Must be run from the directory containing simodel.exe

David Atkins   29-01-00            ver.1.0       */

#include <stdio.h>

#define START_AREA 1
#define STOP_AREA 100
#define STEP_BY 1

int loop_num;

main()
{
  int interaction,district,destination,sub_group;
  char systemcall[200],inString[100],filename[200],group_string[5];
  char home[24]="/home/dja/Analysis/";
  FILE *in_flows,*out_flows;

  /* ALL GLOBAL CALIBRATIONS */

  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Population/population100.csv >
%sModels/CompetingDestinations/cd_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/SocialClass/socialclass100.csv
>> %sModels/CompetingDestinations/cd_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/HousePrices/houseprices100.csv
>> %sModels/CompetingDestinations/cd_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Tenure/tenure100.csv >>
%sModels/CompetingDestinations/cd_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Unemployment/unemployment100.csv
>> %sModels/CompetingDestinations/cd_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat
%sData/ExplanatoryVariables/Accessibility/accessibility100.csv >>
%sModels/CompetingDestinations/cd_tail.txt",home,home);
  system(systemcall);

  sprintf(systemcall,"rm -r %sModels/CompetingDestinations/LatestRun",home);
  system(systemcall);
  sprintf(systemcall,"mkdir %sModels/CompetingDestinations/LatestRun",home);
  system(systemcall);

  /* AGE-GROUP PROGRESSION LOOP */
  for(sub_group=0;sub_group<12;sub_group++)
  {
    if (sub_group == 0)
    {
      strcpy(group_string,"all_1624");
    }
    else if (sub_group == 1)
    {
      strcpy(group_string,"all_2554");
    }
    else if (sub_group == 2)
    {
      strcpy(group_string,"all_55+");
    }
    else if (sub_group == 3)
    {
      strcpy(group_string,"all_16+");
    }
    else if (sub_group == 4)
    {
      strcpy(group_string,"male_1624");
    }
    else if (sub_group == 5)
    {
      strcpy(group_string,"male_2554");
```

```
  }
  else if (sub_group == 6)
  {
    strcpy(group_string,"male_55+");
  }
  else if (sub_group == 7)
  {
    strcpy(group_string,"male_16+");
  }
  else if (sub_group == 8)
  {
    strcpy(group_string,"female_1624");
  }
  else if (sub_group == 9)
  {
    strcpy(group_string,"female_2554");
  }
  else if (sub_group == 10)
  {
    strcpy(group_string,"female_55+");
  }
  else
  {
    strcpy(group_string,"female_16+");
  }

  printf("\nDoing global CompetingDestinations runs\n");
  /* Competing destinations */
  sprintf(systemcall,"mkdir
%sModels/CompetingDestinations/LatestRun/%s",home,group_string);
  system(systemcall);
  sprintf(systemcall,"cat %sModels/CompetingDestinations/cd_head.txt >
%sModels/CompetingDestinations/LatestRun/%s/global.run",home,home,group_string);
  system(systemcall);
  sprintf(systemcall,"cat %sData/FlowData/SMSTable3/%s.txt >>
%sModels/CompetingDestinations/LatestRun/%s/global.run",

home,group_string,home,group_string);
  system(systemcall);
  sprintf(systemcall,"cat %sModels/CompetingDestinations/cd_tail.txt >>
%sModels/CompetingDestinations/LatestRun/%s/global.run",home,home,group_string);
  system(systemcall);

  sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
  system(systemcall);
  sprintf(systemcall,"cp %sModels/CompetingDestinations/LatestRun/%s/global.run
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
  system(systemcall);
  sprintf(systemcall, "%sModels/SIModel/simodel",home);
  system(systemcall);
  sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/CompetingDestinations/LatestRun/%s/global.res",home,home,group_string);
  system(systemcall);

  printf("\nFinished global CompetingDestinations age-gender runs\n");

  /* Perform ORIGIN-SPECIFIC calibrations */
  /* Generate and calibrate run files for OS models */

  printf("\nGenerating OS CompetingDestinations data for %s year old
migrants\n",group_string);
  sprintf(filename,"%sData/FlowData/SMSTable3/%s.txt",home,group_string);
  if ((in_flows=fopen(filename,"r"))==NULL)
  {
    return(printf("Error opening input flow file\n"));
  }
  sprintf(filename,"%sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
  if ((out_flows=fopen(filename,"w"))==NULL)
  {
    return(printf("Error opening output flow file\n"));
  }
  while(fgets(inString,100,in_flows))
  {
    sscanf(inString,"%10i%10i%*5i%5i",&interaction,&district,&destination);
    fprintf(out_flows,"%10i%10i    1%5i\n",interaction,district,destination);
  }
  fclose(in_flows);
  fclose(out_flows);

  /* MAIN ORIGIN PROGRESSION LOOP */
  for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
```

```
    {
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable3/%s.os.txt >
%sData/FlowData/SMSTable3/os.tmp",
            (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/CompetingDestinations/cd_head.os.txt >
%sModels/CompetingDestinations/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable3/os.tmp >>
%sModels/CompetingDestinations/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/CompetingDestinations/cd_tail.txt >>
%sModels/CompetingDestinations/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);

      printf("\nCompeting destinations run for %s year old migrants from origin
%i\n",group_string,loop_num);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/CompetingDestinations/LatestRun/%s/%i.run
%sModels/SIModel/simodel2_0.dat",
              home,group_string,loop_num,home);
      system(systemcall);
      sprintf(systemcall,"%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/CompetingDestinations/LatestRun/%s/%i.res",
              home,home,group_string,loop_num);
      system(systemcall);
    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable3/os.tmp",home);
    system(systemcall);
    sprintf(systemcall,"rm %sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    system(systemcall);
      printf("Completed all calibrations for group: %s/n/n",group_string);
  } /* END OF sub_group PROGRESSION LOOP */

  /* MARITAL STATUS PROGRESSION LOOP */
  for(sub_group=0;sub_group<9;sub_group++)
  {
    if (sub_group == 0)
    {
      strcpy(group_string,"all_single");
    }
    else if (sub_group == 1)
    {
      strcpy(group_string,"all_married");
    }
    else if (sub_group == 2)
    {
      strcpy(group_string,"all_wid_div");
    }
    else if (sub_group == 3)
    {
      strcpy(group_string,"male_single");
    }
    else if (sub_group == 4)
    {
      strcpy(group_string,"male_married");
    }
    else if (sub_group == 5)
    {
      strcpy(group_string,"male_wid_div");
    }
    else if (sub_group == 6)
    {
      strcpy(group_string,"female_single");
    }
    else if (sub_group == 7)
    {
      strcpy(group_string,"female_married");
    }
    else
    {
      strcpy(group_string,"female_wid_div");
    }
```

```
    printf("\nDoing global CompetingDestinations runs\n");
    /* Competing destinations */
    sprintf(systemcall,"mkdir
%sModels/CompetingDestinations/LatestRun/%s",home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/CompetingDestinations/cd_head.txt >
%sModels/CompetingDestinations/LatestRun/%s/trad.run",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable4/%s.txt >>
%sModels/CompetingDestinations/LatestRun/%s/trad.run",

home,group_string,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/CompetingDestinations/cd_tail.txt >>
%sModels/CompetingDestinations/LatestRun/%s/trad.run",home,home,group_string);
    system(systemcall);

    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/CompetingDestinations/LatestRun/%s/trad.run
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/CompetingDestinations/LatestRun/%s/global.res",home,home,group_string);
    system(systemcall);

    printf("\nFinished global CompetingDestinations Marital Status runs\n");

    /* Perform ORIGIN-SPECIFIC calibrations */
    /* Generate and calibrate run files for OS models */

    printf("\nGenerating OS CompetingDestinations data for %s year old
migrants\n",group_string);
    sprintf(filename,"%sData/FlowData/SMSTable4/%s.txt",home,group_string);
    if ((in_flows=fopen(filename,"r"))==NULL)
    {
      return(printf("Error opening input flow file\n"));
    }
    sprintf(filename,"%sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
    if ((out_flows=fopen(filename,"w"))==NULL)
    {
      return(printf("Error opening output flow file\n"));
    }
    while(fgets(inString,100,in_flows))
    {
      sscanf(inString,"%10i%10i%*5i%5i",&interaction,&district,&destination);
      fprintf(out_flows,"%10i%10i    1%5i\n",interaction,district,destination);
    }
    fclose(in_flows);
    fclose(out_flows);

    /* MAIN ORIGIN PROGRESSION LOOP */
    for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
    {
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable4/%s.os.txt >
%sData/FlowData/SMSTable4/os.tmp",
            (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/CompetingDestinations/cd_head.os.txt >
%sModels/CompetingDestinations/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable4/os.tmp >>
%sModels/CompetingDestinations/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/CompetingDestinations/cd_tail.txt >>
%sModels/CompetingDestinations/LatestRun/%s/%i.run",
            home,home,group_string,loop_num);
      system(systemcall);

      printf("\nCompeting destinations run for %s year old migrants from origin
%i\n",group_string,loop_num);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/CompetingDestinations/LatestRun/%s/%i.run
%sModels/SIModel/simodel2_0.dat",
              home,group_string,loop_num,home);
      system(systemcall);
```

```
      sprintf(systemcall,"%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/CompetingDestinations/LatestRun/%s/%i.res",
              home,home,group_string,loop_num);
      system(systemcall);
    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable4/os.tmp",home);
    system(systemcall);
    sprintf(systemcall,"rm %sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
    system(systemcall);

        printf("Completed all calibrations for group: %s/n/n",group_string);
  } /* END OF sub_group PROGRESSION LOOP */

  printf("Finished\n\n");

  return;
}
```

Discrete nested logit model

```
/* This is a C program to automatically generate and run global and origin specific
discrete
nested logit models for 21 migrant subgroups.

Must be run from the directory containing simodel.exe

David Atkins   5Mar2006              ver.2.0        */

#include <stdio.h>

#define START_AREA 50
#define STOP_AREA 50
#define STEP_BY 1
#define debug 0

int loop_num;

main()
{
  long interaction,separation,destination,sub_group, loop1, loop2, loop3, origin_dist;
  long pop[460], hp[460], region[460], sepmat[460][460];
  short district_nl_index[460], pause, diagnostics;
  double socclass[460], tenure[460], unemp[460], dist_pe, pop_pe, socclass_pe, hp_pe,
tenure_pe, unemp_pe, utility[460], rutility[460], sum, mean, min, max, range;
  char systemcall[200],inString[100],filename[200],group_string[5],
tempVarString[100];
  char home[24]="/home/dja/Analysis/";
  FILE *in_flows,*out_flows, *in_var, *in_reg, *out_rutilities;
  void initialiseNLAreas();

  /* Set district_nl_index (which areas to make regionalisations for)      */
  initialiseNLAreas(district_nl_index);

  // Prepare initial tail file (will be added to for second stage calibration
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Population/population100.csv >
%sModels/DiscreteNestedLogit/dnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/SocialClass/socialclass100.csv
>> %sModels/DiscreteNestedLogit/dnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/HousePrices/houseprices100.csv
>> %sModels/DiscreteNestedLogit/dnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Tenure/tenure100.csv >>
%sModels/DiscreteNestedLogit/dnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Unemployment/unemployment100.csv
>> %sModels/DiscreteNestedLogit/dnl_tail.txt",home,home);
  system(systemcall);

  // Read in explanatory variables - need these to calculate the inclusive value for
stage 2 calibration
  sprintf(filename,"%sData/ExplanatoryVariables/Population/population459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%d%*1c",&pop[loop1]);
         if((debug)&&(loop1<5)) printf("pop[%i] = %i\n",loop1,pop[loop1]);
       }
    fclose(in_var);
  }
  else return(0);

  sprintf(filename,"%sData/ExplanatoryVariables/SocialClass/socialclass459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%lf",&socclass[loop1]);
         if((debug)&&(loop1<5)) printf("socclass[%i] = %f\n",loop1,socclass[loop1]);
       }
    fclose(in_var);
  }
  else return(0);

  sprintf(filename,"%sData/ExplanatoryVariables/HousePrices/houseprices459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%d%*1c",&hp[loop1]);
         if((debug)&&(loop1<5)) printf("hp[%i] = %i\n",loop1,hp[loop1]);
       }
```

```
    fclose(in_var);
  }
  else return(0);

  sprintf(filename,"%sData/ExplanatoryVariables/Tenure/tenure459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%lf%*1c",&tenure[loop1]);
          if((debug)&&(loop1<5)) printf("tenure[%i] = %f\n",loop1,tenure[loop1]);
        }
    fclose(in_var);
  }
  else return(0);


sprintf(filename,"%sData/ExplanatoryVariables/Unemployment/unemployment459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%lf%*1c",&unemp[loop1]);
          if((debug)&&(loop1<5)) printf("unemp[%i] = %f\n",loop1,unemp[loop1]);
        }
    fclose(in_var);
  }
  else return(0);

  sprintf(filename,"%sProgs/AccessCalc/sepmat459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      for(loop2=1;loop2<460;loop2++)
        fscanf(in_var,"%d%*1c",&sepmat[loop1][loop2]);
      if((debug)&&(loop1==1)&&(loop2<5)) printf("sepmat[1][%i] =
%i\n",loop1,sepmat[1][loop1]);
        }
    fclose(in_var);
  }
  else return(0);

  printf("Read in explanatory variable data\n");

  // Create LatestRun directory
  sprintf(systemcall,"rm -r %sModels/DiscreteNestedLogit/LatestRun",home);
  system(systemcall);
  sprintf(systemcall,"mkdir %sModels/DiscreteNestedLogit/LatestRun",home);
  system(systemcall);

  // AGE-GROUP PROGRESSION LOOP
  for(sub_group=0;sub_group<12;sub_group++)
  {
    if (sub_group == 0)
      strcpy(group_string,"all_1624");
    else if (sub_group == 1)
      strcpy(group_string,"all_2554");
    else if (sub_group == 2)
      strcpy(group_string,"all_55+");
    else if (sub_group == 3)
      strcpy(group_string,"all_16+");
    else if (sub_group == 4)
      strcpy(group_string,"male_1624");
    else if (sub_group == 5)
      strcpy(group_string,"male_2554");
    else if (sub_group == 6)
      strcpy(group_string,"male_55+");
    else if (sub_group == 7)
      strcpy(group_string,"male_16+");
    else if (sub_group == 8)
      strcpy(group_string,"female_1624");
    else if (sub_group == 9)
      strcpy(group_string,"female_2554");
    else if (sub_group == 10)
      strcpy(group_string,"female_55+");
    else
      strcpy(group_string,"female_16+");


        // Create sub-group output directory within LatestRun directory
        sprintf(systemcall,"mkdir
%sModels/DiscreteNestedLogit/LatestRun/%s",home,group_string);
    system(systemcall);

        // GLOBAL CALIBRATION - using best discrete regionalization calculated for
origin #1 (arbitrary decision - global DNL does really make sense)
```

```
    // Set origin_dist to dist# for nl#1
        for (loop2=1;loop2<460;loop2++)
      if(district_nl_index[loop2] == 1) {
        origin_dist = loop2;
        break;
          }

        // Stage 1 calibration
    printf("\nDoing global stage1 run for %s\n",group_string);
    sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_global_head1.txt >
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable3/%s.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run1",
home,group_string,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_tail.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/LatestRun/%s/global.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/DiscreteNestedLogit/LatestRun/%s/global.res1 ",home,home,group_string);
    system(systemcall);

        // Open stage one results file and extract parameter estimates
    sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
    if((in_var = fopen(filename,"r"))==NULL)
          return(printf("Could open stage 1 results file for %s global DNL
model\n",group_string));
    for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
      fgets(inString, 100, in_var);
    dist_pe=pop_pe=socclass_pe=hp_pe=tenure_pe=unemp_pe=0.0;
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &dist_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &pop_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &socclass_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &hp_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &tenure_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &unemp_pe);
        fclose(in_var);

    /* Read appropriate discrete regionalisation into region[] variable */
    sprintf(filename,"%sProgs/Regionaliser/1.reg1",home);
    in_reg = fopen(filename,"r");
    for(loop2=1;loop2<460;loop2++)
      fscanf(in_reg,"%i%*c",&region[loop2]);
    for(loop2=1;loop2<460;loop2++)
      region[loop2]++;
    fclose(in_reg);
    if(debug)
      for(loop2=1;loop2<10;loop2++)
        printf("district #%i is in region #%i\n",loop2, region[loop2]);


    // Calculate inclusive values based on parameter estimates and district
allocations in discrete regionalization
    utility[origin_dist] = 0.0;
    for(loop2=1;loop2<460;loop2++)
      if(loop2 != origin_dist)
              utility[loop2] = (    dist_pe * log((double)sepmat[origin_dist][loop2])
) +   (pop_pe * log((double)pop[loop2]) ) +
                         (socclass_pe * log(socclass[loop2])          ) +    (hp_pe *
log((double)hp[loop2])  ) +
                         (  tenure_pe * log(tenure[loop2])            ) + (unemp_pe *
log(unemp[loop2])       );
        if (debug)
          for(loop2=1;loop2<460;loop2+=25)
            printf("utility[%i] = %f\n",loop2,utility[loop2]);
```

```
    // Calculate regional inclusive_value[]'s  -  summing exp(utility[]) for ALL
districts in each region
    for(loop2=1;loop2<460;loop2++)
      rutility[loop2]=0.0;
    for(loop2=1;loop2<460;loop2++)
      rutility[region[loop2]]+=(utility[loop2]);

       if(debug)
      for(loop2=1;loop2<10;loop2++)
           printf("rutility[%i] = %f\n",loop2,rutility[loop2]);

    // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
    sum = mean = range = 0.0;
       min = max = rutility[1];
    for(loop2=1;loop2<460;loop2++)
       {
         sum += rutility[loop2];
         if(rutility[loop2]<min) min = rutility[loop2];
         if(rutility[loop2]>max) max = rutility[loop2];
       }
       mean = sum / 459.0;
       range = max - min;
       printf("Mean of regional utility = %f\n",mean);
       printf("Range of regional utility = %f\n",range);
    if(abs(mean)>range)
         for(loop2=1;loop2<460;loop2++)
        rutility[loop2]=rutility[loop2]/abs(mean);
       else
         for(loop2=1;loop2<460;loop2++)
        rutility[loop2]=rutility[loop2]/range;

    // exp() rutility[]
       for(loop2=1;loop2<460;loop2++)
      rutility[loop2]=exp(rutility[loop2]);

    /* Output inclusive values for selected 100 districts (in NL order) to file */
    sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
    system(systemcall);
    sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
    if((out_rutilities = fopen(filename,"w"))==NULL)
         return(printf("Couldn't open %s for writing, exiting...\n",filename));
    for(loop2=1;loop2<100;loop2++)
      for(loop3=1;loop3<460;loop3++)
        if( district_nl_index[loop3] == loop2)
          fprintf(out_rutilities,"%f\n",rutility[region[loop3]]);
    for(loop3=1;loop3<460;loop3++)
      if( district_nl_index[loop3] == 100)
        fprintf(out_rutilities,"%f\n",rutility[region[loop3]]);
        fclose(out_rutilities);

    // SECOND stage calibration
    printf("\nDoing global stage2 run for %s\n",group_string);

    sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/dnl_tail.txt
 %sModels/DiscreteNestedLogit/dnl_tail2.txt",home,home);
    system(systemcall);
    sprintf(systemcall,"cat rutilities.csv >>
 %sModels/DiscreteNestedLogit/dnl_tail2.txt",home);
    system(systemcall);

    sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_global_head2.txt >
 %sModels/DiscreteNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable3/%s.txt >>
 %sModels/DiscreteNestedLogit/LatestRun/%s/global.run2",home,group_string,home,group_st
 ring);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_tail2.txt >>
 %sModels/DiscreteNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/LatestRun/%s/global.run2
 %sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);

    if(debug) {
      sprintf(systemcall,"tail
 %sModels/DiscreteNestedLogit/LatestRun/%s/global.run2",home,group_string);
```

```
      system(systemcall);
        }

    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/DiscreteNestedLogit/LatestRun/%s/global.res ",home,home,group_string);
    system(systemcall);

        printf("\nFinished global DNL run\n");

    /* ORIGIN-SPECIFIC */

    /* Generate 'OS' flow file */
    printf("\nGenerating OS Traditional data for %s year old
migrants\n",group_string);
    sprintf(filename,"%sData/FlowData/SMSTable3/%s.txt",home,group_string);
    if ((in_flows=fopen(filename,"r"))==NULL)
    {
      return(printf("Error opening input flow file\n"));
    }
    sprintf(filename,"%sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    if ((out_flows=fopen(filename,"w"))==NULL)
    {
      return(printf("Error opening output flow file\n"));
    }
    while(fgets(inString,100,in_flows))
    {
      sscanf(inString,"%10i%10i%*5i%5i",&interaction,&separation,&destination);
      fprintf(out_flows,"%10i%10i    1%5i\n",interaction,separation,destination);
    }
    fclose(in_flows);
    fclose(out_flows);

    /* MAIN ORIGIN PROGRESSION LOOP */
    for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
    {
         // Determine which district correlates with loop_num which is an nl#
      for(loop1=0;loop1<460;loop1++)
              if(district_nl_index[loop1]==loop_num)
                origin_dist=loop1;

         // Stage 1 calibration
      printf("\nDoing origin nl#%i stage1 run for %s\n",loop_num,group_string);
      sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_head1.txt >
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable3/%s.os.txt >
%sData/FlowData/SMSTable3/os.tmp", (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable3/os.tmp >>
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run1", home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_tail1.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/LatestRun/%s/%i.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num,home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.res1",home,home,group_string,loop_num);
      system(systemcall);

         // Open stage one results file and extract parameter estimates
      sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
      in_var = fopen(filename,"r");
      for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
        fgets(inString, 100, in_var);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &dist_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &pop_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &socclass_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &hp_pe);
```

```
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%121f", &tenure_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%121f", &unemp_pe);
         fclose(in_var);

      /* Read appropriate discrete regionalisation into region[] variable */
      sprintf(filename,"%sProgs/Regionaliser/%i.regl",home,loop_num);
      in_reg = fopen(filename,"r");
      for(loop2=1;loop2<460;loop2++)
        fscanf(in_reg,"%i%*c",&region[loop2]);
      for(loop2=1;loop2<460;loop2++)
        region[loop2]++;
      fclose(in_reg);

      // Calculate inclusive values based on parameter estimates and district
allocations in discrete regionalization
      utility[origin_dist] = 0.0;
      for(loop2=1;loop2<460;loop2++)
              if(loop2!=origin_dist) {
                utility[loop2] =          (dist_pe *
log((double)sepmat[origin_dist][loop2]) ) +   (pop_pe * log((double)pop[loop2]) ) +
                                        (socclass_pe * log(socclass[loop2])          ) +
(hp_pe * log((double)hp[loop2])  ) +
                                         (tenure_pe * log(tenure[loop2])           ) +
(unemp_pe * log(unemp[loop2])        );
             if((debug)&&(loop2==2))
                {
           printf("OS DNL FOR nl#%i for %s:\n\n",loop_num,group_string);
              printf("PARAMS:\nd = %5.3f\np = %5.3f\ns = %5.3f\nh = %5.3f\nt =
%5.3f\nu = %5.3f\n",dist_pe,pop_pe,socclass_pe,hp_pe,tenure_pe,unemp_pe);
                 printf("VALUES:\nd = %i\np = %i  \ns = %5.2f\nh = %d\nt = %5.2f\nu
=
%5.2f\n",sepmat[1][loop2],pop[loop2],socclass[loop2],hp[loop2],tenure[loop2],unemp[loo
p2]);
                 printf("district %i utility = %8.3f + %8.3f + %8.3f + %8.3f + %8.3f
+ %8.3f = %8.3f\n", loop2,
                                              (dist_pe     *
log((double)sepmat[origin_dist][loop2])),(pop_pe   * log((double)pop[loop2])),
                          (socclass_pe * log(socclass[loop2])),          (hp_pe *
log((double)hp[loop2])),
                          (tenure_pe    * log(tenure[loop2])),         (unemp_pe *
log(unemp[loop2])) ,utility[loop2]);
                }
              }

      // Calculate regional inclusive_value[]'s  -  summing exp(utility) for ALL
districts in each region
        // Should sum ALL districts utilities to get regional utility otherwise the
inclusive values will reflect more the number of
        // NL districts per region that the regional utility as defined by the
discrete regionalization!
      for(loop2=1;loop2<460;loop2++)
        rutility[loop2]=0.0;
      for(loop2=1;loop2<460;loop2++)
        rutility[region[loop2]]+=(utility[loop2]);

      // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
      sum = mean = range = 0.0;
         min = max = rutility[1];
      for(loop2=1;loop2<460;loop2++)
         {
           sum += rutility[loop2];
               if(rutility[loop2]<min) min = rutility[loop2];
               if(rutility[loop2]>max) max = rutility[loop2];
         }
         mean = sum / 459.0;
         range = max - min;
         printf("Mean of regional utility = %f\n",mean);
         printf("Range of regional utility = %f\n",range);
      if(abs(mean)>range)
           for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/abs(mean);
         else
           for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/range;

         // exp() rutility[]
         for(loop2=1;loop2<460;loop2++) {
        rutility[loop2]=exp(rutility[loop2]);
```

```
            if(rutility[loop2]>1000.0) {
              printf("rutility[%i]=%f\nThis is a suspiciously high regional utility
value, outputing diagnostics:\n",loop2,rutility[loop2]);
              diagnostics = 1;
              pause = 1;
            }
        }

        if(diagnostics) {
          for(loop3=1;loop3<460;loop3++) {
          printf("OS DNL FOR nl#%i for %s:\n",loop_num,group_string);
             printf("PARAMS:
%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\n",dist_pe,pop_pe,socclass_pe,hp_pe,tenure_pe
,unemp_pe);
                 printf("VALUES:
%i\t%i\t%5.2f\t%d\t%5.2f\t%5.2f\n",sepmat[1][loop3],pop[loop3],socclass[loop3],hp[loop
3],tenure[loop3],unemp[loop3]);
                 printf("district %i utility = %8.3f + %8.3f + %8.3f + %8.3f + %8.3f +
%8.3f = %8.3f\n", loop3,
                                            (dist_pe     *
log((double)sepmat[origin_dist][loop3])),(pop_pe   * log((double)pop[loop3])),
                        (socclass_pe * log(socclass[loop3])),           (hp_pe *
log((double)hp[loop3])),
                        (tenure_pe   * log(tenure[loop3])),          (unemp_pe *
log(unemp[loop3])) ,utility[loop3]);
                }
        printf("\n");
          }

      /* Output inclusive values for selected 100 districts (in NL order) to file */
      sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
      system(systemcall);
      sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
      if((out_rutilities = fopen(filename,"w"))==NULL)
           return(printf("Couldn't open %s for writing, exiting...\n",filename));
      for(loop2=1;loop2<100;loop2++)
        for(loop3=1;loop3<460;loop3++)
          if( district_nl_index[loop3] == loop2)
            fprintf(out_rutilities,"%.3f\n",rutility[region[loop3]]);
      for(loop3=1;loop3<460;loop3++)
        if( district_nl_index[loop3] == 100)
          fprintf(out_rutilities,"%.3f\n",rutility[region[loop3]]);
      fclose(out_rutilities);

      // SECOND stage calibration
      printf("\nDoing origin nl#%i stage2 run for %s\n",loop_num,group_string);

      sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/dnl_tail.txt
%sModels/DiscreteNestedLogit/dnl_tail2.txt",home,home);
      system(systemcall);
      sprintf(systemcall,"cat rutilities.csv >>
%sModels/DiscreteNestedLogit/dnl_tail2.txt",home);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_head2.txt >
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable3/%s.os.txt >
%sData/FlowData/SMSTable3/os.tmp", (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable3/os.tmp >>
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run2", home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_tail2.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/LatestRun/%s/%i.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num, home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.res ",home,home,group_string,loop_num);
      system(systemcall);

      printf("\nFinished origin nl#%i DNL run\n", loop_num);

          if((debug)&&(loop_num<4)) {
                printf("In debug stepping mode... press any key to
continue...\n",loop_num,group_string);
```

```
                getchar();
        }

        if(pause) {
          printf("Pausing due to output of diagnostic info, press any key to
continue...\n");
          getchar();
          pause = 0;
              diagnostics = 0;
        }

    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    system(systemcall);
        printf("Completed all calibrations for group: %s\n\n",group_string);
  } /* END OF sub_group PROGRESSION LOOP */

  // MARITAL STATUS PROGRESSION LOOP
  for(sub_group=0;sub_group<9;sub_group++)
  {
    if (sub_group == 0)
      strcpy(group_string,"all_single");
    else if (sub_group == 1)
      strcpy(group_string,"all_married");
    else if (sub_group == 2)
      strcpy(group_string,"all_wid_div");
    else if (sub_group == 3)
      strcpy(group_string,"male_single");
    else if (sub_group == 4)
      strcpy(group_string,"male_married");
    else if (sub_group == 5)
      strcpy(group_string,"male_wid_div");
    else if (sub_group == 6)
      strcpy(group_string,"female_single");
    else if (sub_group == 7)
      strcpy(group_string,"female_married");
    else
      strcpy(group_string,"female_wid_div");

        // Create sub-group output directory within LatestRun directory
        sprintf(systemcall,"mkdir
%sModels/DiscreteNestedLogit/LatestRun/%s",home,group_string);
    system(systemcall);

        // GLOBAL CALIBRATION - using best discrete regionalization calculated for
origin #1 (arbitrary decision - global DNL does really make sense)

    // Set origin_dist to dist# for nl#1
        for (loop2=1;loop2<460;loop2++)
      if(district_nl_index[loop2] == 1) {
        origin_dist = loop2;
        break;
         }

        // Stage 1 calibration
    printf("\nDoing global stage1 run for %s\n",group_string);
    sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_global_head1.txt >
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable4/%s.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run1",home,group_string,home,group_st
ring);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_tail.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/LatestRun/%s/global.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/DiscreteNestedLogit/LatestRun/%s/global.res1 ",home,home,group_string);
    system(systemcall);

        // Open stage one results file and extract parameter estimates
    sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
    in_var = fopen(filename,"r");
```

```
    for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
      fgets(inString, 100, in_var);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &dist_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &pop_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &socclass_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &hp_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &tenure_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &unemp_pe);
    fclose(in_var);
    /* Read appropriate discrete regionalisation into region[] variable */
    sprintf(filename,"%sProgs/Regionaliser/1.regl",home);
    in_reg = fopen(filename,"r");
    for(loop2=1;loop2<460;loop2++)
      fscanf(in_reg,"%i%*c",&region[loop2]);
    for(loop2=1;loop2<460;loop2++)
      region[loop2]++;
    fclose(in_reg);

    // Calculate inclusive values based on parameter estimates and district
allocations in discrete regionalization
    utility[origin_dist] = 0.0;
    for(loop2=1;loop2<460;loop2++)
         if(loop2 != origin_dist) {
           utility[loop2] = (    dist_pe * log((double)sepmat[origin_dist][loop2]) ) +
(pop_pe * log((double)pop[loop2]) ) +
                                (socclass_pe * log(socclass[loop2])          ) +    (hp_pe
* log((double)hp[loop2])  ) +
                                (  tenure_pe * log(tenure[loop2])            ) + (unemp_pe
* log(unemp[loop2])         );
           if((debug)&&(loop2==2))
              {
          printf("GLOBAL DNL FOR %s:\n\n",group_string);

             printf("PARAMS:\nd = %5.3f\np = %5.3f\ns = %5.3f\nh = %5.3f\nt = %5.3f\nu
= %5.3f\n",
                                                        dist_pe,     pop_pe,
socclass_pe,hp_pe,           tenure_pe,     unemp_pe);

                printf("VALUES:\nd = %i\np = %i  \ns = %5.2f\nh = %d\nt = %5.2f\nu =
%5.2f\n",
                                          sepmat[1][loop2],  pop[loop2],  socclass[loop2],
hp[loop2],  tenure[loop2],  unemp[loop2]);

                printf("district %i utility = %8.3f + %8.3f + %8.3f + %8.3f + %8.3f +
%8.3f = %8.3f\n", loop2,
                                             (dist_pe     *
log((double)sepmat[origin_dist][loop2])),(pop_pe   * log((double)pop[loop2])),
                              (socclass_pe * log(socclass[loop2])),          (hp_pe *
log((double)hp[loop2])),
                              (tenure_pe    * log(tenure[loop2])),          (unemp_pe *
log(unemp[loop2])) ,utility[loop2]);
              }
         }

     // Calculate regional inclusive_value[]'s  -  summing exp(utility) for ALL
districts in each region
     for(loop2=1;loop2<460;loop2++)
       rutility[loop2]=0.0;
     for(loop2=1;loop2<460;loop2++)
       rutility[region[loop2]]+=(utility[loop2]);

     // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
     sum = mean = range = 0.0;
        min = max = rutility[1];
     for(loop2=1;loop2<460;loop2++)
        {
          sum += rutility[loop2];
          if(rutility[loop2]<min) min = rutility[loop2];
          if(rutility[loop2]>max) max = rutility[loop2];
        }
        mean = sum / 459.0;
        range = max - min;
        printf("Mean of regional utility = %f\n",mean);
```

```
      printf("Range of regional utility = %f\n",range);
   if(abs(mean)>range)
        for(loop2=1;loop2<460;loop2++)
       rutility[loop2]=rutility[loop2]/abs(mean);
      else
        for(loop2=1;loop2<460;loop2++)
       rutility[loop2]=rutility[loop2]/range;

   // exp() rutility[]
      for(loop2=1;loop2<460;loop2++)
     rutility[loop2]=exp(rutility[loop2]);

   /* Output inclusive values for selected 100 districts (in NL order) to file */
   sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
   system(systemcall);
   sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
   if((out_rutilities = fopen(filename,"w"))==NULL)
        return(printf("Couldn't open %s for writing, exiting...\n",filename));
   for(loop2=1;loop2<100;loop2++)
     for(loop3=1;loop3<460;loop3++)
       if( district_nl_index[loop3] == loop2)
         fprintf(out_rutilities,"%.3f\n",rutility[region[loop3]]);
   for(loop3=1;loop3<460;loop3++)
     if( district_nl_index[loop3] == 100)
       fprintf(out_rutilities,"%.3f\n",rutility[region[loop3]]);
       fclose(out_rutilities);

   // SECOND stage calibration
   printf("\nDoing global stage2 run for %s\n",group_string);

   sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/dnl_tail.txt
%sModels/DiscreteNestedLogit/dnl_tail2.txt",home,home);
   system(systemcall);
   sprintf(systemcall,"cat rutilities.csv >>
%sModels/DiscreteNestedLogit/dnl_tail2.txt",home);
   system(systemcall);
   sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_global_head2.txt >
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
   system(systemcall);
   sprintf(systemcall,"cat %sData/FlowData/SMSTable4/%s.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run2",home,group_string,home,group_st
ring);
   system(systemcall);
   sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_tail2.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
   system(systemcall);
   sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
   system(systemcall);
   sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/LatestRun/%s/global.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
   system(systemcall);
   sprintf(systemcall, "%sModels/SIModel/simodel",home);
   system(systemcall);
   sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/DiscreteNestedLogit/LatestRun/%s/global.res ",home,home,group_string);
   system(systemcall);

       printf("\nFinished global DNL run\n");

    /* ORIGIN-SPECIFIC */

    /* Generate 'OS' flow file */
    printf("\nGenerating OS Traditional data for %s year old
migrants\n",group_string);
    sprintf(filename,"%sData/FlowData/SMSTable4/%s.txt",home,group_string);
    if ((in_flows=fopen(filename,"r"))==NULL)
    {
      return(printf("Error opening input flow file\n"));
    }
    sprintf(filename,"%sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
    if ((out_flows=fopen(filename,"w"))==NULL)
    {
      return(printf("Error opening output flow file\n"));
    }
    while(fgets(inString,100,in_flows))
    {
      sscanf(inString,"%10i%10i%*5i%5i",&interaction,&separation,&destination);
      fprintf(out_flows,"%10i%10i    1%5i\n",interaction,separation,destination);
    }
    fclose(in_flows);
    fclose(out_flows);
```

```
    /* MAIN ORIGIN PROGRESSION LOOP */
    for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
    {
         // Determine which district correlates with loop_num which is an nl#
      for(loop1=0;loop1<460;loop1++)
              if(district_nl_index[loop1]==loop_num)
                origin_dist=loop1;

         // Stage 1 calibration
      printf("\nDoing origin nl#%i stage1 run for %s\n",loop_num,group_string);
      sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_head1.txt >
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable4/%s.os.txt >
%sData/FlowData/SMSTable4/os.tmp",(loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable4/os.tmp >>
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_tail.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/LatestRun/%s/%i.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num,home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.res1",home,home,group_string,loop_num);
      system(systemcall);

         // Open stage one results file and extract parameter estimates
      sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
      in_var = fopen(filename,"r");
      for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
        fgets(inString, 100, in_var);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &dist_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &pop_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &socclass_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &hp_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &tenure_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &unemp_pe);
         fclose(in_var);

      /* Read appropriate discrete regionalisation into region[] variable */
      sprintf(filename,"%sProgs/Regionaliser/%i.reg1",home,loop_num);
      in_reg = fopen(filename,"r");
      for(loop2=1;loop2<460;loop2++)
        fscanf(in_reg,"%i%*c",&region[loop2]);
      for(loop2=1;loop2<460;loop2++)
        region[loop2]++;
      fclose(in_reg);

      // Calculate inclusive values based on parameter estimates and district
allocations in discrete regionalization
      utility[origin_dist] = 0.0;
      for(loop2=1;loop2<460;loop2++)
        if(loop2!=origin_dist)
              utility[loop2] =        (dist_pe *
log((double)sepmat[origin_dist][loop2]) ) +   (pop_pe * log((double)pop[loop2]) ) +
                              (socclass_pe * log(socclass[loop2])          ) +
(hp_pe * log((double)hp[loop2])  ) +
                                (tenure_pe * log(tenure[loop2])            ) +
(unemp_pe * log(unemp[loop2])        );

      // Calculate regional inclusive_value[]'s  -  summing exp(utility) for ALL
districts in each region
          // Should sum ALL districts utilities to get regional utility otherwise the
inclusive values will reflect more the number of
          // NL districts per region that the regional utility as defined by the
discrete regionalization!
```

```
      for(loop2=1;loop2<460;loop2++)
        rutility[loop2]=0.0;
      for(loop2=1;loop2<460;loop2++)
        rutility[region[loop2]]+=utility[loop2];

      // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
      sum = mean = range = 0.0;
          min = max = rutility[1];
      for(loop2=1;loop2<460;loop2++)
          {
            sum += rutility[loop2];
                if(rutility[loop2]<min) min = rutility[loop2];
                if(rutility[loop2]>max) max = rutility[loop2];
          }
          mean = sum / 459.0;
          range = max - min;
          printf("Mean of regional utility = %f\n",mean);
          printf("Range of regional utility = %f\n",range);
      if(abs(mean)>range)
            for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/abs(mean);
          else
            for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/range;

          // exp() rutility[]
          for(loop2=1;loop2<460;loop2++) {
        rutility[loop2]=exp(rutility[loop2]);
                if(rutility[loop2]>1000.0) {
                  printf("rutility[%i]=%f\nThis is a suspiciously high regional utility
value, outputing diagnostics:\n",loop2,rutility[loop2]);
                  diagnostics = 1;
                  pause = 1;
                }
          }

          if(diagnostics) {
            for(loop3=1;loop3<460;loop3++) {
          printf("OS DNL FOR nl#%i for %s:\n",loop_num,group_string);
              printf("PARAMS:
%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\n",dist_pe,pop_pe,socclass_pe,hp_pe,tenure_pe
,unemp_pe);
                  printf("VALUES:
%i\t%i\t%5.2f\t%d\t%5.2f\t%5.2f\n",sepmat[1][loop3],pop[loop3],socclass[loop3],hp[loop
3],tenure[loop3],unemp[loop3]);
                  printf("district %i utility = %8.3f + %8.3f + %8.3f + %8.3f + %8.3f +
%8.3f = %8.3f\n", loop3,
                                                    (dist_pe     *
log((double)sepmat[origin_dist][loop3])),(pop_pe    * log((double)pop[loop3])),
                              (socclass_pe * log(socclass[loop3])),             (hp_pe *
log((double)hp[loop3])),
                              (tenure_pe    * log(tenure[loop3])),            (unemp_pe *
log(unemp[loop3])) ,utility[loop3]);
                }
        printf("\n");
          }

       /* Output inclusive values for selected 100 districts (in NL order) to file */
       sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
       system(systemcall);
       sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
       if((out_rutilities = fopen(filename,"w"))==NULL)
            return(printf("Couldn't open %s for writing, exiting...\n",filename));
       for(loop2=1;loop2<100;loop2++)
         for(loop3=1;loop3<460;loop3++)
           if( district_nl_index[loop3] == loop2)
             fprintf(out_rutilities,"%.3f\n",rutility[region[loop3]]);
       for(loop3=1;loop3<460;loop3++)
         if( district_nl_index[loop3] == 100)
           fprintf(out_rutilities,"%.3f\n",rutility[region[loop3]]);
          fclose(out_rutilities);

       // SECOND stage calibration
       printf("\nDoing origin nl#%i stage2 run for %s\n",loop_num,group_string);

       sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/dnl_tail.txt
 %sModels/DiscreteNestedLogit/dnl_tail2.txt",home,home);
       system(systemcall);
       sprintf(systemcall,"cat rutilities.csv >>
 %sModels/DiscreteNestedLogit/dnl_tail2.txt",home);
```

```
      system(systemcall);
      sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_head2.txt >
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable4/%s.os.txt >
%sData/FlowData/SMSTable4/os.tmp", (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable4/os.tmp >>
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run2", home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/DiscreteNestedLogit/dnl_tail2.txt >>
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/DiscreteNestedLogit/LatestRun/%s/%i.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num, home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/DiscreteNestedLogit/LatestRun/%s/%i.res ",home,home,group_string,loop_num);
      system(systemcall);

      printf("\nFinished origin nl#%i DNL run\n", loop_num);

          if((debug)&&(loop_num<4)) {
                 printf("In debug stepping mode... press any key to
continue...\n",loop_num,group_string);
                 getchar();
          }

          if(pause) {
            printf("Pausing due to output of diagnostic info, press any key to
continue...\n");
            getchar();
            pause = 0;
                diagnostics = 0;
          }

    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
    system(systemcall);
        printf("Completed all calibrations for group: %s\n\n",group_string);
  } /* END OF sub_group PROGRESSION LOOP */

  printf("Finished all DNL calibration runs!\n\n");

  return;
}

void initialiseNLAreas(nl_areas)
short *nl_areas;
{
  int loop;
  for(loop=0;loop<460;loop++)
    nl_areas[loop]=0;

  nl_areas[418]=1;
  nl_areas[15]=2;
  nl_areas[49]=3;
  nl_areas[70]=4;
  nl_areas[58]=5;
  nl_areas[241]=6;
  nl_areas[242]=7;
  nl_areas[34]=8;
  nl_areas[141]=9;
  nl_areas[65]=10;
  nl_areas[157]=11;
  nl_areas[71]=12;
  nl_areas[35]=13;
  nl_areas[91]=14;
  nl_areas[2]=15;
  nl_areas[228]=16;
  nl_areas[398]=17;
  nl_areas[118]=18;
  nl_areas[168]=19;
  nl_areas[178]=20;
  nl_areas[97]=21;
  nl_areas[169]=22;
```

```
    nl_areas[59]=23;
    nl_areas[20]=24;
    nl_areas[150]=25;
    nl_areas[125]=26;
    nl_areas[50]=27;
    nl_areas[230]=28;
    nl_areas[60]=29;
    nl_areas[455]=30;
    nl_areas[152]=31;
    nl_areas[21]=32;
    nl_areas[432]=33;
    nl_areas[132]=34;
    nl_areas[445]=35;
    nl_areas[181]=36;
    nl_areas[23]=37;
    nl_areas[341]=38;
    nl_areas[3]=39;
    nl_areas[4]=40;
    nl_areas[293]=41;
    nl_areas[24]=42;
    nl_areas[334]=43;
    nl_areas[6]=44;
    nl_areas[7]=45;
    nl_areas[274]=46;
    nl_areas[223]=47;
    nl_areas[8]=48;
    nl_areas[247]=49;
    nl_areas[68]=50;
    nl_areas[259]=51;
    nl_areas[266]=52;
    nl_areas[45]=53;
    nl_areas[76]=54;
    nl_areas[102]=55;
    nl_areas[233]=56;
    nl_areas[36]=57;
    nl_areas[107]=58;
    nl_areas[88]=59;
    nl_areas[81]=60;
    nl_areas[54]=61;
    nl_areas[382]=62;
    nl_areas[282]=63;
    nl_areas[276]=64;
    nl_areas[305]=65;
    nl_areas[37]=66;
    nl_areas[308]=67;
    nl_areas[95]=68;
    nl_areas[135]=69;
    nl_areas[145]=70;
    nl_areas[192]=71;
    nl_areas[249]=72;
    nl_areas[82]=73;
    nl_areas[38]=74;
    nl_areas[234]=75;
    nl_areas[51]=76;
    nl_areas[39]=77;
    nl_areas[296]=78;
    nl_areas[52]=79;
    nl_areas[194]=80;
    nl_areas[11]=81;
    nl_areas[211]=82;
    nl_areas[328]=83;
    nl_areas[410]=84;
    nl_areas[40]=85;
    nl_areas[330]=86;
    nl_areas[353]=87;
    nl_areas[57]=88;
    nl_areas[403]=89;
    nl_areas[365]=90;
    nl_areas[12]=91;
    nl_areas[42]=92;
    nl_areas[69]=93;
    nl_areas[63]=94;
    nl_areas[104]=95;
    nl_areas[354]=96;
    nl_areas[43]=97;
    nl_areas[85]=98;
    nl_areas[64]=99;
    nl_areas[298]=100;
}
```

Weighted nested logit model

```
/* This is a C program to automatically generate and run global and origin specific
weighted nested logit models for 21 migrant subgroups.

Must be run from the directory containing simodel.exe

David Atkins   5Mar2006              ver.2.0        */

#include <stdio.h>

#define START_AREA 50
#define STOP_AREA 50
#define STEP_BY 1
#define debug 0

int loop_num;
short district_nl_index[460];

main()
{
  long interaction,separation,destination,sub_group, loop1, loop2, loop3, origin_dist;
  long pop[460], hp[460], regionalization[460][460], sepmat[460][460],
reg_row_totals[460];
  short pause, diagnostics;
  double socclass[460], tenure[460], unemp[460], dist_pe, pop_pe, socclass_pe, hp_pe,
tenure_pe, unemp_pe, utility[460], rutility[460], sum, mean, min, max, range;
  char systemcall[200],inString[100],filename[200],group_string[5],
tempVarString[100];
  char home[24]="/home/dja/Analysis/";
  FILE *in_flows,*out_flows, *in_var, *in_reg, *out_rutilities;
  void initialiseNLAreas();

  /* Set district_nl_index (which areas to make regionalisations for)      */
  initialiseNLAreas();

  pause = 0; // Used to pause processing if diagnostic information is output

  // Prepare initial tail file (will be added to for second stage calibration
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Population/population100.csv >
%sModels/WeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/SocialClass/socialclass100.csv
>> %sModels/WeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/HousePrices/houseprices100.csv
>> %sModels/WeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Tenure/tenure100.csv >>
%sModels/WeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Unemployment/unemployment100.csv
>> %sModels/WeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);

  // Read in explanatory variables - need these to calculate the inclusive value for
stage 2 calibration
  sprintf(filename,"%sData/ExplanatoryVariables/Population/population459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%d%*1c",&pop[loop1]);
        if((debug)&&(loop1<5)) printf("pop[%i] = %i\n",loop1,pop[loop1]);
      }
    fclose(in_var);
  }
  else return(0);

  sprintf(filename,"%sData/ExplanatoryVariables/SocialClass/socialclass459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%lf",&socclass[loop1]);
        if(debug) printf("socclass[%i] = %f\n",loop1,socclass[loop1]);
      }
    fclose(in_var);
  }
  else return(0);

  sprintf(filename,"%sData/ExplanatoryVariables/HousePrices/houseprices459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
```

```
    fscanf(in_var,"%d%*1c",&hp[loop1]);
      if(debug) printf("hp[%i] = %i\n",loop1,hp[loop1]);
    }
  fclose(in_var);
}
else return(0);

sprintf(filename,"%sData/ExplanatoryVariables/Tenure/tenure459.csv",home);
if((in_var = fopen(filename,"r"))!=NULL) {
  for(loop1=1;loop1<460;loop1++) {
    fscanf(in_var,"%lf%*1c",&tenure[loop1]);
      if(debug) printf("tenure[%i] = %f\n",loop1,tenure[loop1]);
    }
  fclose(in_var);
}
else return(0);


sprintf(filename,"%sData/ExplanatoryVariables/Unemployment/unemployment459.csv",home);
if((in_var = fopen(filename,"r"))!=NULL) {
  for(loop1=1;loop1<460;loop1++) {
    fscanf(in_var,"%lf%*1c",&unemp[loop1]);
      if(debug) printf("unemp[%i] = %f\n",loop1,unemp[loop1]);
    }
  fclose(in_var);
}
else return(0);

sprintf(filename,"%sProgs/AccessCalc/sepmat459.csv",home);
if((in_var = fopen(filename,"r"))!=NULL) {
  for(loop1=1;loop1<460;loop1++) {
       for(loop2=1;loop2<460;loop2++) {
      fscanf(in_var,"%d%*1c",&sepmat[loop1][loop2]);
      if((debug)&&(loop1==1)) printf("sepmat[1][%i] = %i\n",loop2,sepmat[1][loop2]);
       }
    }
  fclose(in_var);
}
else return(0);

printf("Read in explanatory variable data\n");

// Create LatestRun directory
sprintf(systemcall,"rm -r %sModels/WeightedNestedLogit/LatestRun",home);
system(systemcall);
sprintf(systemcall,"mkdir %sModels/WeightedNestedLogit/LatestRun",home);
system(systemcall);

// AGE-GROUP PROGRESSION LOOP
for(sub_group=0;sub_group<12;sub_group++)
{
  if (sub_group == 0)
    strcpy(group_string,"all_1624");
  else if (sub_group == 1)
    strcpy(group_string,"all_2554");
  else if (sub_group == 2)
    strcpy(group_string,"all_55+");
  else if (sub_group == 3)
    strcpy(group_string,"all_16+");
  else if (sub_group == 4)
    strcpy(group_string,"male_1624");
  else if (sub_group == 5)
    strcpy(group_string,"male_2554");
  else if (sub_group == 6)
    strcpy(group_string,"male_55+");
  else if (sub_group == 7)
    strcpy(group_string,"male_16+");
  else if (sub_group == 8)
    strcpy(group_string,"female_1624");
  else if (sub_group == 9)
    strcpy(group_string,"female_2554");
  else if (sub_group == 10)
    strcpy(group_string,"female_55+");
  else
    strcpy(group_string,"female_16+");

      if(debug) {
        sub_group = 3;
        strcpy(group_string,"all_16+");
      }
```

```
        // Create sub-group output directory within LatestRun directory
        sprintf(systemcall,"mkdir
%sModels/WeightedNestedLogit/LatestRun/%s",home,group_string);
    system(systemcall);

        // GLOBAL CALIBRATION - using best Weighted regionalization calculated for
origin #1 (arbitrary decision - global WNL does really make sense)

    // Set origin_dist to dist# for nl#1
        for (loop2=1;loop2<460;loop2++)
      if(district_nl_index[loop2] == 1) {
        origin_dist = loop2;
        break;
          }
        printf("Origin_dist# set to %i\n",origin_dist);

        // Stage 1 calibration
    printf("\nDoing global stage1 run for %s\n",group_string);
    sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_global_head1.txt >
%sModels/WeightedNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable3/%s.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/global.run1",
home,group_string,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_tail.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/WeightedNestedLogit/LatestRun/%s/global.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/WeightedNestedLogit/LatestRun/%s/global.res1 ",home,home,group_string);
    system(systemcall);

        // Open stage one results file and extract parameter estimates
    sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
    if((in_var = fopen(filename,"r"))==NULL)
         return(printf("Could not open stage 1 results file for %s global WNL
model\n",group_string));
    for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
      fgets(inString, 100, in_var);
    dist_pe = pop_pe = socclass_pe = hp_pe = tenure_pe = unemp_pe = 99999.0;
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &dist_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &pop_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &socclass_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &hp_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &tenure_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &unemp_pe);
      fclose(in_var);
    printf("Global ParaEsts: %f\t%f\t%f\t%f\t%f\t%f\n",dist_pe,pop_pe, socclass_pe,
hp_pe, tenure_pe, unemp_pe);

    // Read appropriate Weighted regionalisation into regionalization[][] variable
    sprintf(filename,"%sProgs/Regionaliser/1.matrix",home);
    in_reg = fopen(filename,"r");
    for(loop2=1;loop2<460;loop2++)
         for(loop3=1;loop3<460;loop3++)
       fscanf(in_reg,"%i%*c",&regionalization[loop2][loop3]);
    fclose(in_reg);

    // Calculate utility of each individual district
    utility[origin_dist] = 0.0;
    for(loop2=1;loop2<460;loop2++)
         if(loop2 != origin_dist)
           utility[loop2] = (    dist_pe * log((double)sepmat[origin_dist][loop2]) ) +
(pop_pe * log((double)pop[loop2]) ) +
                              (socclass_pe * log(socclass[loop2])          ) +    (hp_pe
* log((double)hp[loop2])  ) +
```

```
                               (  tenure_pe * log(tenure[loop2])          ) + (unemp_pe
* log(unemp[loop2])        );

    // Calculate inclusive_value[]'s  -  summing weighted exp(utility) of ALL
districts
    for(loop2=1;loop2<460;loop2++)
      rutility[loop2]=0.0;
    for(loop2=1;loop2<460;loop2++)
      for(loop3=1;loop3<460;loop3++)

rutility[loop2]+=(utility[loop3]*(double)(regionalization[loop2][loop3])/1000.0);

    // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
    sum = mean = range = 0.0;
    min = max = rutility[1];
    for(loop2=1;loop2<460;loop2++)
        {
          sum += rutility[loop2];
          if(rutility[loop2]<min) min = rutility[loop2];
      if(rutility[loop2]>max) max = rutility[loop2];
        }
        mean = sum / 459.0;
        range = max - min;
        printf("Mean of regional utility = %f\n",mean);
        printf("Range of regional utility = %f\n",range);
    if(abs(mean)>range)
          for(loop2=1;loop2<460;loop2++)
         rutility[loop2]=rutility[loop2]/abs(mean);
        else
          for(loop2=1;loop2<460;loop2++)
         rutility[loop2]=rutility[loop2]/range;

        // Apply exp() to rutility[]
        for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=exp(rutility[loop2]);

    /* Output inclusive values for selected 100 districts (in NL order) to file */
    sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
    system(systemcall);
    sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
    if((out_rutilities = fopen(filename,"w"))==NULL)
          return(printf("Couldn't open %s for writing, exiting...\n",filename));
    for(loop2=1;loop2<101;loop2++)
      for(loop3=1;loop3<460;loop3++)
        if( district_nl_index[loop3] == loop2)
          fprintf(out_rutilities,"%f\n",rutility[loop3]);
       fclose(out_rutilities);

    // SECOND stage calibration
    printf("\nDoing global stage2 run for %s\n",group_string);

    sprintf(systemcall,"cp %sModels/WeightedNestedLogit/wnl_tail.txt
 %sModels/WeightedNestedLogit/wnl_tail2.txt",home,home);
    system(systemcall);
    sprintf(systemcall,"cat rutilities.csv >>
 %sModels/WeightedNestedLogit/wnl_tail2.txt",home);
    system(systemcall);

    sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_global_head2.txt >
 %sModels/WeightedNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable3/%s.txt >>
 %sModels/WeightedNestedLogit/LatestRun/%s/global.run2",home,group_string,home,group_st
 ring);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_tail2.txt >>
 %sModels/WeightedNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/WeightedNestedLogit/LatestRun/%s/global.run2
 %sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);

    if(debug) {
      sprintf(systemcall,"tail
 %sModels/WeightedNestedLogit/LatestRun/%s/global.run2",home,group_string);
      system(systemcall);
       }
```

```
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/WeightedNestedLogit/LatestRun/%s/global.res ",home,home,group_string);
    system(systemcall);

        printf("\nFinished global WNL run\n");

    /* ORIGIN-SPECIFIC */

    /* Generate 'OS' flow file */
    printf("\nGenerating OS Traditional data for %s year old
migrants\n",group_string);
    sprintf(filename,"%sData/FlowData/SMSTable3/%s.txt",home,group_string);
    if ((in_flows=fopen(filename,"r"))==NULL)
    {
      return(printf("Error opening input flow file\n"));
    }
    sprintf(filename,"%sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    if ((out_flows=fopen(filename,"w"))==NULL)
    {
      return(printf("Error opening output flow file\n"));
    }
    while(fgets(inString,100,in_flows))
    {
      sscanf(inString,"%10i%10i%*5i%5i",&interaction,&separation,&destination);
      fprintf(out_flows,"%10i%10i    1%5i\n",interaction,separation,destination);
    }
    fclose(in_flows);
    fclose(out_flows);

    /* MAIN ORIGIN PROGRESSION LOOP */
    for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
    {
          // Determine which district correlates with loop_num which is an nl#
      for(loop1=0;loop1<460;loop1++)
              if(district_nl_index[loop1]==loop_num)
                origin_dist=loop1;
          printf("\nOrigin_dist# set to %i\n",origin_dist);

          // Stage 1 calibration
      printf("Doing origin nl#%i stage1 run for %s\n",loop_num,group_string);
      sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_head1.txt >
 %sModels/WeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable3/%s.os.txt >
 %sData/FlowData/SMSTable3/os.tmp", (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable3/os.tmp >>
 %sModels/WeightedNestedLogit/LatestRun/%s/%i.run1", home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_tail.txt >>
 %sModels/WeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/WeightedNestedLogit/LatestRun/%s/%i.run1
 %sModels/SIModel/simodel2_0.dat",home,group_string,loop_num,home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
 %sModels/WeightedNestedLogit/LatestRun/%s/%i.res1",home,home,group_string,loop_num);
      system(systemcall);

          // Open stage one results file and extract parameter estimates
      sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
      in_var = fopen(filename,"r");
      dist_pe = pop_pe = socclass_pe = hp_pe = tenure_pe = unemp_pe = 99999.0;
      for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
 file
          fgets(inString, 100, in_var);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &dist_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &pop_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &socclass_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%12lf", &hp_pe);
      fgets(inString, 100, in_var);
```

```
sscanf(inString,"%*20c%12lf", &tenure_pe);
fgets(inString, 100, in_var);
sscanf(inString,"%*20c%12lf", &unemp_pe);
    fclose(in_var);

// Read appropriate Weighted regionalisation into regionalization[][] variable
sprintf(filename,"%sProgs/Regionaliser/%i.matrix",home,loop_num);
in_reg = fopen(filename,"r");
for(loop2=1;loop2<460;loop2++)
     for(loop3=1;loop3<460;loop3++)
    fscanf(in_reg,"%i%*c",&regionalization[loop2][loop3]);
fclose(in_reg);
if(debug)
    {
     printf("Weighted Regionalization Matrix:-\n");
  for(loop2=1;loop2<5;loop2++)
         {
       for(loop3=1;loop3<6;loop3++)
      printf("%i,",regionalization[loop2][loop3]);
    printf("%i\n",regionalization[loop2][5]);
         }
         printf("\n");
    }

// Calculate utility of each individual district
utility[origin_dist] = 0.0;
for(loop2=1;loop2<460;loop2++)
        if(loop2!=origin_dist)
          utility[loop2] = (    dist_pe *
log((double)sepmat[origin_dist][loop2]) ) +    (pop_pe * log((double)pop[loop2]) ) +
                                  (socclass_pe * log(socclass[loop2])              ) +
(hp_pe * log((double)hp[loop2])  ) +
                                  (  tenure_pe * log(tenure[loop2])                ) +
(unemp_pe * log(unemp[loop2])        );

if(debug && loop_num<5)
  for(loop2=1;loop2<11;loop2++)
    for(loop3=1;loop3<460;loop3++)
      if(district_nl_index[loop3]==loop2) {
        printf("ParaEstimates    : %f\t%f\t%f\t%f\t%f\t%f\n",dist_pe,pop_pe,
socclass_pe, hp_pe, tenure_pe, unemp_pe);
        printf("Variables[nl#%i]:
%f\t%f\t%f\t%f\t%f\t%f\n",loop2,(double)sepmat[origin_dist][loop3],(double)pop[loop3],
socclass[loop3],(double)hp[loop3],tenure[loop3],unemp[loop3]);
            printf("utility[nl#%i]    : %f\n\n",loop2,utility[loop3]);
                 }

// Calculate inclusive_value[]'s  -  summing weighted exp(utility) of ALL
districts
for(loop2=1;loop2<460;loop2++)
  rutility[loop2]=0.0;
for(loop2=1;loop2<460;loop2++)
  for(loop3=1;loop3<460;loop3++)

rutility[loop2]+=(utility[loop3]*(double)(regionalization[loop2][loop3])/1000.0);
if(debug && loop_num<5)
  for(loop2=1;loop2<11;loop2++)
    for(loop3=1;loop3<460;loop3++)
      if(district_nl_index[loop3]==loop2)
           printf("rutility[nl%i] = %f\n",loop2,rutility[loop3]);

// Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
sum = mean = range = 0.0;
    min = max = rutility[1];
for(loop2=1;loop2<460;loop2++)
    {
      sum += rutility[loop2];
           if(rutility[loop2]<min) min = rutility[loop2];
           if(rutility[loop2]>max) max = rutility[loop2];
    }
    mean = sum / 459.0;
    range = max - min;
    printf("Mean of regional utility = %f\n",mean);
    printf("Range of regional utility = %f\n",range);
if(abs(mean)>range)
      for(loop2=1;loop2<460;loop2++)
     rutility[loop2]=rutility[loop2]/abs(mean);
    else
      for(loop2=1;loop2<460;loop2++)
     rutility[loop2]=rutility[loop2]/range;
```

```
    if(debug && loop_num<5)
      for(loop2=1;loop2<11;loop2++)
        for(loop3=1;loop3<460;loop3++)
          if(district_nl_index[loop3]==loop2)
               printf("normalized rutility[nl%i] = %f\n",loop2,rutility[loop3]);

        // exp() rutility[]
        for(loop2=1;loop2<460;loop2++) {
      rutility[loop2]=exp(rutility[loop2]);
             if(abs(rutility[loop2])>1000.0) {
               printf("rutility[%i]=%f\nThis is a suspiciously high regional utility
value, outputing diagnostics:\n",loop2,rutility[loop2]);
               diagnostics = 1;
               pause = 1;
             }
        }

        if(diagnostics) {
          for(loop3=1;loop3<460;loop3++) {
        printf("OS DNL FOR nl#%i for %s:\n",loop_num,group_string);
            printf("PARAMS:
%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\n",dist_pe,pop_pe,socclass_pe,hp_pe,tenure_pe
,unemp_pe);
               printf("VALUES:
%i\t%i\t%5.2f\t%d\t%5.2f\t%5.2f\n",sepmat[1][loop3],pop[loop3],socclass[loop3],hp[loop
3],tenure[loop3],unemp[loop3]);
               printf("district %i utility = %8.3f + %8.3f + %8.3f + %8.3f + %8.3f +
%8.3f = %8.3f\n", loop3,
                                            (dist_pe     *
log((double)sepmat[origin_dist][loop3])),(pop_pe   * log((double)pop[loop3])),
                          (socclass_pe * log(socclass[loop3])),          (hp_pe *
log((double)hp[loop3])),
                          (tenure_pe   * log(tenure[loop3])),          (unemp_pe *
log(unemp[loop3])) ,utility[loop3]);
             }
       printf("\n");
        }

     if(debug && loop_num<5)
       for(loop2=1;loop2<11;loop2++)
         for(loop3=1;loop3<460;loop3++)
           if(district_nl_index[loop3]==loop2)
                printf("exp'd normalized rutility[nl%i] =
%f\n",loop2,rutility[loop3]);

     // Output inclusive values for selected 100 districts (in NL order) to file
     sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
     system(systemcall);
     sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
     if((out_rutilities = fopen(filename,"w"))==NULL)
          return(printf("Couldn't open %s for writing, exiting...\n",filename));
     for(loop2=1;loop2<101;loop2++)
       for(loop3=1;loop3<460;loop3++)
         if( district_nl_index[loop3] == loop2)
           fprintf(out_rutilities,"%f\n",rutility[loop3]);
     fclose(out_rutilities);

     // SECOND stage calibration
     printf("\nDoing origin nl#%i stage2 run for %s\n",loop_num,group_string);

     sprintf(systemcall,"cp %sModels/WeightedNestedLogit/wnl_tail.txt
%sModels/WeightedNestedLogit/wnl_tail2.txt",home,home);
     system(systemcall);
     sprintf(systemcall,"cat rutilities.csv >>
%sModels/WeightedNestedLogit/wnl_tail2.txt",home);
     system(systemcall);
     sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_head2.txt >
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_num);
     system(systemcall);
     sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable3/%s.os.txt >
%sData/FlowData/SMSTable3/os.tmp", (loop_num-1)*100+1,home,group_string,home);
     system(systemcall);
     sprintf(systemcall,"head -100 %sData/FlowData/SMSTable3/os.tmp >>
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run2", home,home,group_string,loop_num);
     system(systemcall);
     sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_tail2.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_num);
     system(systemcall);
     sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
     system(systemcall);
```

```
      sprintf(systemcall,"cp %sModels/WeightedNestedLogit/LatestRun/%s/%i.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num, home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/WeightedNestedLogit/LatestRun/%s/%i.res ",home,home,group_string,loop_num);
      system(systemcall);

      printf("\nFinished origin nl#%i WNL run\n", loop_num);

         if((debug)&&(loop_num<4)) {
                printf("In debug stepping mode... press any key to
continue...\n",loop_num,group_string);
                getchar();
         }

         if(pause) {
           printf("Pausing due to output of diagnostic info, press any key to
continue...\n");
           getchar();
           pause = 0;
               diagnostics = 0;
         }

    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    system(systemcall);
        printf("Completed all calibrations for group: %s\n\n",group_string);

    if(debug)
          return(printf("Terminating after first set of global + OS runs because
running in debug mode\n"));

  } /* END OF sub_group PROGRESSION LOOP */

  // AGE-GROUP PROGRESSION LOOP
  for(sub_group=0;sub_group<9;sub_group++)
  {
    if (sub_group == 0)
      strcpy(group_string,"all_single");
    else if (sub_group == 1)
      strcpy(group_string,"all_married");
    else if (sub_group == 2)
      strcpy(group_string,"all_wid_div");
    else if (sub_group == 3)
      strcpy(group_string,"male_single");
    else if (sub_group == 4)
      strcpy(group_string,"male_married");
    else if (sub_group == 5)
      strcpy(group_string,"male_wid_div");
    else if (sub_group == 6)
      strcpy(group_string,"female_single");
    else if (sub_group == 7)
      strcpy(group_string,"female_married");
    else
      strcpy(group_string,"female_wid_div");

        // Create sub-group output directory within LatestRun directory
        sprintf(systemcall,"mkdir
%sModels/WeightedNestedLogit/LatestRun/%s",home,group_string);
    system(systemcall);

        // GLOBAL CALIBRATION - using best Weighted regionalization calculated for
origin #1 (arbitrary decision - global WNL does really make sense)

    // Set origin_dist to dist# for nl#1
        for (loop2=1;loop2<460;loop2++)
      if(district_nl_index[loop2] == 1) {
        origin_dist = loop2;
        break;
          }
        printf("Origin_dist# set to %i\n",origin_dist);

        // Stage 1 calibration
    printf("\nDoing global stage1 run for %s\n",group_string);
    sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_global_head1.txt >
%sModels/WeightedNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
```

```
    sprintf(systemcall,"cat %sData/FlowData/SMSTable4/%s.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/global.run1",home,group_string,home,group_st
ring);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_tail.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/WeightedNestedLogit/LatestRun/%s/global.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/WeightedNestedLogit/LatestRun/%s/global.res1 ",home,home,group_string);
    system(systemcall);

        // Open stage one results file and extract parameter estimates
    sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
    in_var = fopen(filename,"r");
    dist_pe = pop_pe = socclass_pe = hp_pe = tenure_pe = unemp_pe = 99999.0;
    for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
      fgets(inString, 100, in_var);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &dist_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &pop_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &socclass_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &hp_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &tenure_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &unemp_pe);
    fclose(in_var);

    // Read appropriate Weighted regionalisation into regionalization[][] variable
    sprintf(filename,"%sProgs/Regionaliser/1.matrix",home);
    in_reg = fopen(filename,"r");
    for(loop2=1;loop2<460;loop2++)
        for(loop3=1;loop3<460;loop3++)
      fscanf(in_reg,"%i%*c",&regionalization[loop2][loop3]);
    fclose(in_reg);
    if(debug)
      {
        printf("Weighted Regionalization Matrix:-\n");
     for(loop2=1;loop2<5;loop2++)
        {
          for(loop3=1;loop3<5;loop3++)
         printf("%i,",regionalization[loop2][loop3]);
       printf("%i\n",regionalization[loop2][5]);
        }
      }

    // Calculate utility of each individual district
    utility[origin_dist] = 0.0;
    for(loop2=1;loop2<460;loop2++)
        if (loop2 != origin_dist) {
          utility[loop2] = (    dist_pe * log((double)sepmat[origin_dist][loop2]) ) +
(pop_pe * log((double)pop[loop2]) ) +
                         (socclass_pe * log(socclass[loop2])          ) +    (hp_pe
* log((double)hp[loop2])  ) +
                         (  tenure_pe * log(tenure[loop2])            ) + (unemp_pe
* log(unemp[loop2])       );
        }

    if (debug)
        for(loop2=1;loop2<460;loop2+=25)
          printf("utility[%i] = %f\n",loop2,utility[loop2]);

    // Calculate inclusive_value[]'s  -  summing weighted exp(utility) of ALL
districts
    for(loop2=1;loop2<460;loop2++)
      rutility[loop2]=0.0;
    for(loop2=1;loop2<460;loop2++)
      for(loop3=1;loop3<460;loop3++)

rutility[loop2]+=(utility[loop3]*(double)(regionalization[loop2][loop3])/1000.0);
```

```
if(debug)
 for(loop2=1;loop2<5;loop2++)
      printf("rutility[%i] = %f\n",loop2,rutility[loop2]);

  // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
  sum = mean = range = 0.0;
  min = max = rutility[1];
  for(loop2=1;loop2<460;loop2++)
    {
      sum += rutility[loop2];
      if(rutility[loop2]<min) min = rutility[loop2];
   if(rutility[loop2]>max) max = rutility[loop2];
    }
    mean = sum / 459.0;
    range = max - min;
    printf("Mean of regional utility = %f\n",mean);
    printf("Range of regional utility = %f\n",range);
  if(abs(mean)>range)
      for(loop2=1;loop2<460;loop2++)
     rutility[loop2]=rutility[loop2]/abs(mean);
    else
      for(loop2=1;loop2<460;loop2++)
     rutility[loop2]=rutility[loop2]/range;

    // Apply exp() to rutility[]
    for(loop2=1;loop2<460;loop2++)
      rutility[loop2]=exp(rutility[loop2]);

  /* Output inclusive values for selected 100 districts (in NL order) to file */
  sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
  system(systemcall);
  sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
  if((out_rutilities = fopen(filename,"w"))==NULL)
       return(printf("Couldn't open %s for writing, exiting...\n",filename));
  for(loop2=1;loop2<101;loop2++)
    for(loop3=1;loop3<460;loop3++)
      if( district_nl_index[loop3] == loop2)
        fprintf(out_rutilities,"%f\n",rutility[loop3]);
     fclose(out_rutilities);

  // SECOND stage calibration
  printf("\nDoing global stage2 run for %s\n",group_string);

  sprintf(systemcall,"cp %sModels/WeightedNestedLogit/wnl_tail.txt
%sModels/WeightedNestedLogit/wnl_tail2.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat rutilities.csv >>
%sModels/WeightedNestedLogit/wnl_tail2.txt",home);
  system(systemcall);
  sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_global_head2.txt >
%sModels/WeightedNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
  system(systemcall);
  sprintf(systemcall,"cat %sData/FlowData/SMSTable4/%s.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/global.run2",home,group_string,home,group_st
ring);
  system(systemcall);
  sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_tail2.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
  system(systemcall);
  sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
  system(systemcall);
  sprintf(systemcall,"cp %sModels/WeightedNestedLogit/LatestRun/%s/global.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
  system(systemcall);
  sprintf(systemcall, "%sModels/SIModel/simodel",home);
  system(systemcall);
  sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/WeightedNestedLogit/LatestRun/%s/global.res ",home,home,group_string);
  system(systemcall);

     printf("\nFinished global WNL run\n");

  /* ORIGIN-SPECIFIC */

  /* Generate 'OS' flow file */
  printf("\nGenerating OS Traditional data for %s year old
migrants\n",group_string);
  sprintf(filename,"%sData/FlowData/SMSTable4/%s.txt",home,group_string);
  if ((in_flows=fopen(filename,"r"))==NULL)
  {
```

```
    return(printf("Error opening input flow file\n"));
  }
  sprintf(filename,"%sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
  if ((out_flows=fopen(filename,"w"))==NULL)
  {
    return(printf("Error opening output flow file\n"));
  }
  while(fgets(inString,100,in_flows))
  {
    sscanf(inString,"%10i%10i%*5i%5i",&interaction,&separation,&destination);
    fprintf(out_flows,"%10i%10i    1%5i\n",interaction,separation,destination);
  }
  fclose(in_flows);
  fclose(out_flows);

  /* MAIN ORIGIN PROGRESSION LOOP */
  for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
  {
       // Determine which district correlates with loop_num which is an nl#
    for(loop1=0;loop1<460;loop1++)
            if(district_nl_index[loop1]==loop_num)
              origin_dist=loop1;
       printf("\nOrigin_dist# set to %i\n",origin_dist);

       // Stage 1 calibration
    printf("Doing origin nl#%i stage1 run for %s\n",loop_num,group_string);
    sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_head1.txt >
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
    system(systemcall);
    sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable4/%s.os.txt >
%sData/FlowData/SMSTable4/os.tmp",(loop_num-1)*100+1,home,group_string,home);
    system(systemcall);
    sprintf(systemcall,"head -100 %sData/FlowData/SMSTable4/os.tmp >>
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_tail.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_num);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/WeightedNestedLogit/LatestRun/%s/%i.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/WeightedNestedLogit/LatestRun/%s/%i.res1",home,home,group_string,loop_num);
    system(systemcall);

        // Open stage one results file and extract parameter estimates
    sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
    in_var = fopen(filename,"r");
    dist_pe = pop_pe = socclass_pe = hp_pe = tenure_pe = unemp_pe = 99999.0;
    for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
       fgets(inString, 100, in_var);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &dist_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &pop_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &socclass_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &hp_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &tenure_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &unemp_pe);
        fclose(in_var);

    // Read appropriate Weighted regionalisation into regionalization[][] variable
    sprintf(filename,"%sProgs/Regionaliser/%i.matrix",home,loop_num);
    in_reg = fopen(filename,"r");
    for(loop2=1;loop2<460;loop2++)
         for(loop3=1;loop3<460;loop3++)
       fscanf(in_reg,"%i%*c",&regionalization[loop2][loop3]);
    fclose(in_reg);
    if(debug)
       {
         printf("Weighted Regionalization Matrix:-\n");
      for(loop2=1;loop2<5;loop2++)
```

```
            {
           for(loop3=1;loop3<5;loop3++)
          printf("%i,",regionalization[loop2][loop3]);
         printf("%i\n",regionalization[loop2][5]);
            }
        }

      // Calculate utility of each individual district
      utility[origin_dist] = 0.0;
      for(loop2=1;loop2<460;loop2++)
        if(loop2!=origin_dist)
                utility[loop2] = (    dist_pe *
log((double)sepmat[origin_dist][loop2]) ) +   (pop_pe * log((double)pop[loop2]) ) +
                              (socclass_pe * log(socclass[loop2])           )
+    (hp_pe * log((double)hp[loop2])  ) +
                              (  tenure_pe * log(tenure[loop2])             )
+ (unemp_pe * log(unemp[loop2])        );
      if (debug)
           for(loop2=1;loop2<460;loop2+=25)
             printf("utility[%i] = %f\n",loop2,utility[loop2]);

      // Calculate inclusive_value[]'s  -  summing weighted exp(utility) of ALL
districts
      for(loop2=1;loop2<460;loop2++)
        rutility[loop2]=0.0;
      for(loop2=1;loop2<460;loop2++)
        for(loop3=1;loop3<460;loop3++)

rutility[loop2]+=(utility[loop3]*(double)(regionalization[loop2][loop3])/1000.0);

      if(debug)
        for(loop2=1;loop2<5;loop2++)
             printf("rutility[%i] = %f\n",loop2,rutility[loop2]);

      // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
      sum = mean = range = 0.0;
         min = max = rutility[1];
      for(loop2=1;loop2<460;loop2++)
         {
           sum += rutility[loop2];
               if(rutility[loop2]<min) min = rutility[loop2];
               if(rutility[loop2]>max) max = rutility[loop2];
         }
         mean = sum / 459.0;
         range = max - min;
         printf("Mean of regional utility = %f\n",mean);
         printf("Range of regional utility = %f\n",range);
      if(abs(mean)>range)
           for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/abs(mean);
         else
           for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/range;

                      // exp() rutility[]
         for(loop2=1;loop2<460;loop2++) {
       rutility[loop2]=exp(rutility[loop2]);
              if(abs(rutility[loop2])>1000.0) {
                printf("rutility[%i]=%f\nThis is a suspiciously high regional utility
value, outputing diagnostics:\n",loop2,rutility[loop2]);
                diagnostics = 1;
                pause = 1;
              }
         }

         if(diagnostics) {
           for(loop3=1;loop3<460;loop3++) {
          printf("OS DNL FOR nl#%i for %s:\n",loop_num,group_string);
              printf("PARAMS:
%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\n",dist_pe,pop_pe,socclass_pe,hp_pe,tenure_pe
,unemp_pe);
                printf("VALUES:
%i\t%i\t%5.2f\t%d\t%5.2f\t%5.2f\n",sepmat[1][loop3],pop[loop3],socclass[loop3],hp[loop
3],tenure[loop3],unemp[loop3]);
                printf("district %i utility = %8.3f + %8.3f + %8.3f + %8.3f + %8.3f +
%8.3f = %8.3f\n", loop3,
                                             (dist_pe      *
log((double)sepmat[origin_dist][loop3])),(pop_pe    * log((double)pop[loop3])),
                          (socclass_pe * log(socclass[loop3])),            (hp_pe *
log((double)hp[loop3])),
```

```
                              (tenure_pe   * log(tenure[loop3])),          (unemp_pe *
log(unemp[loop3])) ,utility[loop3]);
                    }
        printf("\n");
          }


          /* Output inclusive values for selected 100 districts (in NL order) to file
*/
      sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
      system(systemcall);
      sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
      if((out_rutilities = fopen(filename,"w"))==NULL)
            return(printf("Couldn't open %s for writing, exiting...\n",filename));
      for(loop2=1;loop2<101;loop2++)
        for(loop3=1;loop3<460;loop3++)
          if( district_nl_index[loop3] == loop2)
            fprintf(out_rutilities,"%f\n",rutility[loop3]);
          fclose(out_rutilities);

      // SECOND stage calibration
      printf("\nDoing origin nl#%i stage2 run for %s\n",loop_num,group_string);

      sprintf(systemcall,"cp %sModels/WeightedNestedLogit/wnl_tail.txt
%sModels/WeightedNestedLogit/wnl_tail2.txt",home,home);
      system(systemcall);
      sprintf(systemcall,"cat rutilities.csv >>
%sModels/WeightedNestedLogit/wnl_tail2.txt",home);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_head2.txt >
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable4/%s.os.txt >
%sData/FlowData/SMSTable4/os.tmp", (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable4/os.tmp >>
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run2", home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/WeightedNestedLogit/wnl_tail2.txt >>
%sModels/WeightedNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/WeightedNestedLogit/LatestRun/%s/%i.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num, home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/WeightedNestedLogit/LatestRun/%s/%i.res ",home,home,group_string,loop_num);
      system(systemcall);

      printf("\nFinished origin nl#%i WNL run\n", loop_num);

          if((debug)&&(loop_num<4)) {
                  printf("In debug stepping mode... press any key to
continue...\n",loop_num,group_string);
                  getchar();
          }

          if(pause) {
            printf("Pausing due to output of diagnostic info, press any key to
continue...\n");
            getchar();
            pause = 0;
                diagnostics = 0;
          }

    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
    system(systemcall);
        printf("Completed all calibrations for group: %s\n\n",group_string);
  } /* END OF sub_group PROGRESSION LOOP */

  printf("Finished all WNL calibration runs!\n\n");

  return;
}

void initialiseNLAreas()
```

```
{
  int loop;
  for(loop=1;loop<460;loop++)
    district_nl_index[loop]=0;

  district_nl_index[418]=1;
  district_nl_index[15]=2;
  district_nl_index[49]=3;
  district_nl_index[70]=4;
  district_nl_index[58]=5;
  district_nl_index[241]=6;
  district_nl_index[242]=7;
  district_nl_index[34]=8;
  district_nl_index[141]=9;
  district_nl_index[65]=10;
  district_nl_index[157]=11;
  district_nl_index[71]=12;
  district_nl_index[35]=13;
  district_nl_index[91]=14;
  district_nl_index[2]=15;
  district_nl_index[228]=16;
  district_nl_index[398]=17;
  district_nl_index[118]=18;
  district_nl_index[168]=19;
  district_nl_index[178]=20;
  district_nl_index[97]=21;
  district_nl_index[169]=22;
  district_nl_index[59]=23;
  district_nl_index[20]=24;
  district_nl_index[150]=25;
  district_nl_index[125]=26;
  district_nl_index[50]=27;
  district_nl_index[230]=28;
  district_nl_index[60]=29;
  district_nl_index[455]=30;
  district_nl_index[152]=31;
  district_nl_index[21]=32;
  district_nl_index[432]=33;
  district_nl_index[132]=34;
  district_nl_index[445]=35;
  district_nl_index[181]=36;
  district_nl_index[23]=37;
  district_nl_index[341]=38;
  district_nl_index[3]=39;
  district_nl_index[4]=40;
  district_nl_index[293]=41;
  district_nl_index[24]=42;
  district_nl_index[334]=43;
  district_nl_index[6]=44;
  district_nl_index[7]=45;
  district_nl_index[274]=46;
  district_nl_index[223]=47;
  district_nl_index[8]=48;
  district_nl_index[247]=49;
  district_nl_index[68]=50;
  district_nl_index[259]=51;
  district_nl_index[266]=52;
  district_nl_index[45]=53;
  district_nl_index[76]=54;
  district_nl_index[102]=55;
  district_nl_index[233]=56;
  district_nl_index[36]=57;
  district_nl_index[107]=58;
  district_nl_index[88]=59;
  district_nl_index[81]=60;
  district_nl_index[54]=61;
  district_nl_index[382]=62;
  district_nl_index[282]=63;
  district_nl_index[276]=64;
  district_nl_index[305]=65;
  district_nl_index[37]=66;
  district_nl_index[308]=67;
  district_nl_index[95]=68;
  district_nl_index[135]=69;
  district_nl_index[145]=70;
  district_nl_index[192]=71;
  district_nl_index[249]=72;
  district_nl_index[82]=73;
  district_nl_index[38]=74;
  district_nl_index[234]=75;
  district_nl_index[51]=76;
```

```
    district_nl_index[39]=77;
    district_nl_index[296]=78;
    district_nl_index[52]=79;
    district_nl_index[194]=80;
    district_nl_index[11]=81;
    district_nl_index[211]=82;
    district_nl_index[328]=83;
    district_nl_index[410]=84;
    district_nl_index[40]=85;
    district_nl_index[330]=86;
    district_nl_index[353]=87;
    district_nl_index[57]=88;
    district_nl_index[403]=89;
    district_nl_index[365]=90;
    district_nl_index[12]=91;
    district_nl_index[42]=92;
    district_nl_index[69]=93;
    district_nl_index[63]=94;
    district_nl_index[104]=95;
    district_nl_index[354]=96;
    district_nl_index[43]=97;
    district_nl_index[85]=98;
    district_nl_index[64]=99;
    district_nl_index[298]=100;
}
```

Hybrid weighted nested logit model

```
/* This is a C program to automatically generate and run global and origin specific
weighted nested logit models for 21 migrant subgroups.

Must be run from the directory containing simodel.exe

David Atkins   5Mar2006             ver.2.0        */

#include <stdio.h>

#define START_AREA 1
#define STOP_AREA 100
#define STEP_BY 1
#define debug 0

int loop_num;
short district_nl_index[460];

main()
{
  long interaction,separation,destination,sub_group, loop1, loop2, loop3, origin_dist;
  long pop[460], hp[460], regionalization[460][460], sepmat[460][460],
reg_row_totals[460];
  short pause, diagnostics;
  double socclass[460], tenure[460], unemp[460], access[460], dist_pe, pop_pe,
socclass_pe, hp_pe, tenure_pe, unemp_pe, access_pe;
  double utility[460], rutility[460], sum, mean, min, max, range;
  char systemcall[200],inString[100],filename[200],group_string[5],
tempVarString[100];
  char home[24]="/home/dja/Analysis/";
  FILE *in_flows,*out_flows, *in_var, *in_reg, *out_rutilities;
  void initialiseNLAreas();

  /* Set district_nl_index (which areas to make regionalisations for)       */
  initialiseNLAreas();

  pause = 0; // Used to pause processing if diagnostic information is output

  // Prepare initial tail file (will be added to for second stage calibration
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Population/population100.csv >
%sModels/HybridWeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/SocialClass/socialclass100.csv
>> %sModels/HybridWeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/HousePrices/houseprices100.csv
>> %sModels/HybridWeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Tenure/tenure100.csv >>
%sModels/HybridWeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat %sData/ExplanatoryVariables/Unemployment/unemployment100.csv
>> %sModels/HybridWeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);
  sprintf(systemcall,"cat
%sData/ExplanatoryVariables/Accessibility/accessibility100.csv >>
%sModels/HybridWeightedNestedLogit/wnl_tail.txt",home,home);
  system(systemcall);

  // Read in explanatory variables - need these to calculate the inclusive value for
stage 2 calibration
  sprintf(filename,"%sData/ExplanatoryVariables/Population/population459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%d%*1c",&pop[loop1]);
          if((debug)&&(loop1<5)) printf("pop[%i] = %i\n",loop1,pop[loop1]);
        }
    fclose(in_var);
  }
  else return(0);

  sprintf(filename,"%sData/ExplanatoryVariables/SocialClass/socialclass459.csv",home);
  if((in_var = fopen(filename,"r"))!=NULL) {
    for(loop1=1;loop1<460;loop1++) {
      fscanf(in_var,"%lf",&socclass[loop1]);
          if(debug) printf("socclass[%i] = %f\n",loop1,socclass[loop1]);
        }
    fclose(in_var);
  }
```

```
else return(0);

sprintf(filename,"%sData/ExplanatoryVariables/HousePrices/houseprices459.csv",home);
if((in_var = fopen(filename,"r"))!=NULL) {
  for(loop1=1;loop1<460;loop1++) {
    fscanf(in_var,"%d%*1c",&hp[loop1]);
       if(debug) printf("hp[%i] = %i\n",loop1,hp[loop1]);
     }
  fclose(in_var);
}
else return(0);

sprintf(filename,"%sData/ExplanatoryVariables/Tenure/tenure459.csv",home);
if((in_var = fopen(filename,"r"))!=NULL) {
  for(loop1=1;loop1<460;loop1++) {
    fscanf(in_var,"%lf%*1c",&tenure[loop1]);
       if(debug) printf("tenure[%i] = %f\n",loop1,tenure[loop1]);
     }
  fclose(in_var);
}
else return(0);


sprintf(filename,"%sData/ExplanatoryVariables/Unemployment/unemployment459.csv",home);
 if((in_var = fopen(filename,"r"))!=NULL) {
   for(loop1=1;loop1<460;loop1++) {
     fscanf(in_var,"%lf%*1c",&unemp[loop1]);
        if(debug) printf("unemp[%i] = %f\n",loop1,unemp[loop1]);
      }
   fclose(in_var);
 }
 else return(0);


sprintf(filename,"%sData/ExplanatoryVariables/Accessibility/accessibility459.csv",home
);
 if((in_var = fopen(filename,"r"))!=NULL) {
   for(loop1=1;loop1<460;loop1++) {
     fscanf(in_var,"%lf%*1c",&access[loop1]);
        if(debug) printf("access[%i] = %f\n",loop1,unemp[loop1]);
      }
   fclose(in_var);
 }
 else return(0);

 sprintf(filename,"%sProgs/AccessCalc/sepmat459.csv",home);
 if((in_var = fopen(filename,"r"))!=NULL) {
   for(loop1=1;loop1<460;loop1++) {
         for(loop2=1;loop2<460;loop2++) {
       fscanf(in_var,"%d%*1c",&sepmat[loop1][loop2]);
       if((debug)&&(loop1==1)) printf("sepmat[1][%i] = %i\n",loop2,sepmat[1][loop2]);
         }
       }
   fclose(in_var);
 }
 else return(0);

 printf("Read in explanatory variable data\n");

 // Create LatestRun directory
 sprintf(systemcall,"rm -r %sModels/HybridWeightedNestedLogit/LatestRun",home);
 system(systemcall);
 sprintf(systemcall,"mkdir %sModels/HybridWeightedNestedLogit/LatestRun",home);
 system(systemcall);

 // AGE-GROUP PROGRESSION LOOP
 for(sub_group=0;sub_group<12;sub_group++)
 {
   if (sub_group == 0)
     strcpy(group_string,"all_1624");
   else if (sub_group == 1)
     strcpy(group_string,"all_2554");
   else if (sub_group == 2)
     strcpy(group_string,"all_55+");
   else if (sub_group == 3)
     strcpy(group_string,"all_16+");
   else if (sub_group == 4)
     strcpy(group_string,"male_1624");
   else if (sub_group == 5)
     strcpy(group_string,"male_2554");
   else if (sub_group == 6)
```

```
      strcpy(group_string,"male_55+");
    else if (sub_group == 7)
      strcpy(group_string,"male_16+");
    else if (sub_group == 8)
      strcpy(group_string,"female_1624");
    else if (sub_group == 9)
      strcpy(group_string,"female_2554");
    else if (sub_group == 10)
      strcpy(group_string,"female_55+");
    else
      strcpy(group_string,"female_16+");

        if(debug) {
          sub_group = 3;
          strcpy(group_string,"all_16+");
        }

        // Create sub-group output directory within LatestRun directory
        sprintf(systemcall,"mkdir
%sModels/HybridWeightedNestedLogit/LatestRun/%s",home,group_string);
    system(systemcall);

        // GLOBAL CALIBRATION - using best Weighted regionalization calculated for
origin #1 (arbitrary decision - global WNL does really make sense)

    // Set origin_dist to dist# for nl#1
        for (loop2=1;loop2<460;loop2++)
      if(district_nl_index[loop2] == 1) {
        origin_dist = loop2;
        break;
          }
        printf("Origin_dist# set to %i\n",origin_dist);

        // Stage 1 calibration
    printf("\nDoing global stage1 run for %s\n",group_string);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_global_head1.txt >
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable3/%s.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run1",
home,group_string,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_tail.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.res1 ",home,home,group_string);
    system(systemcall);

        // Open stage one results file and extract parameter estimates
    sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
    if((in_var = fopen(filename,"r"))==NULL)
          return(printf("Could not open stage 1 results file for %s global WNL
model\n",group_string));
    for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
        fgets(inString, 100, in_var);
    dist_pe = pop_pe = socclass_pe = hp_pe = tenure_pe = unemp_pe = 99999.0;
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &dist_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &pop_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &socclass_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &hp_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &tenure_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%12lf", &unemp_pe);
        fclose(in_var);
    sscanf(inString,"%*20c%12lf", &access_pe);
        fclose(in_var);
```

```
    printf("Global ParaEsts: %f\t%f\t%f\t%f\t%f\t%f\t%f\n",dist_pe,pop_pe,
socclass_pe, hp_pe, tenure_pe, unemp_pe, access_pe);

    // Read appropriate Weighted regionalisation into regionalization[][] variable
    sprintf(filename,"%sProgs/Regionaliser/1.matrix",home);
    in_reg = fopen(filename,"r");
    for(loop2=1;loop2<460;loop2++)
         for(loop3=1;loop3<460;loop3++)
       fscanf(in_reg,"%i%*c",&regionalization[loop2][loop3]);
    fclose(in_reg);

    // Calculate utility of each individual district
    utility[origin_dist] = 0.0;
    for(loop2=1;loop2<460;loop2++)
         if(loop2 != origin_dist)
           utility[loop2] = (    dist_pe * log((double)sepmat[origin_dist][loop2]) ) +
(pop_pe * log((double)pop[loop2]) ) +
                             (socclass_pe * log(socclass[loop2])          ) +    (hp_pe
* log((double)hp[loop2])  ) +
                             (  tenure_pe * log(tenure[loop2])            ) + (unemp_pe
* log(unemp[loop2])       ) +
                                                  (  access_pe * log(access[loop2])
);

    // Calculate inclusive_value[]'s  -  summing weighted exp(utility) of ALL
districts
    for(loop2=1;loop2<460;loop2++)
      rutility[loop2]=0.0;
    for(loop2=1;loop2<460;loop2++)
      for(loop3=1;loop3<460;loop3++)

rutility[loop2]+=(utility[loop3]*(double)(regionalization[loop2][loop3])/1000.0);

    // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
    sum = mean = range = 0.0;
    min = max = rutility[1];
    for(loop2=1;loop2<460;loop2++)
        {
          sum += rutility[loop2];
          if(rutility[loop2]<min) min = rutility[loop2];
      if(rutility[loop2]>max) max = rutility[loop2];
        }
        mean = sum / 459.0;
        range = max - min;
        printf("Mean of regional utility = %f\n",mean);
        printf("Range of regional utility = %f\n",range);
    if(abs(mean)>range)
          for(loop2=1;loop2<460;loop2++)
         rutility[loop2]=rutility[loop2]/abs(mean);
        else
          for(loop2=1;loop2<460;loop2++)
         rutility[loop2]=rutility[loop2]/range;

        // Apply exp() to rutility[]
        for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=exp(rutility[loop2]);

    /* Output inclusive values for selected 100 districts (in NL order) to file */
    sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
    system(systemcall);
    sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
    if((out_rutilities = fopen(filename,"w"))==NULL)
          return(printf("Couldn't open %s for writing, exiting...\n",filename));
    for(loop2=1;loop2<101;loop2++)
      for(loop3=1;loop3<460;loop3++)
        if( district_nl_index[loop3] == loop2)
          fprintf(out_rutilities,"%f\n",rutility[loop3]);
        fclose(out_rutilities);

    // SECOND stage calibration
    printf("\nDoing global stage2 run for %s\n",group_string);

    sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/wnl_tail.txt
%sModels/HybridWeightedNestedLogit/wnl_tail2.txt",home,home);
    system(systemcall);
    sprintf(systemcall,"cat rutilities.csv >>
%sModels/HybridWeightedNestedLogit/wnl_tail2.txt",home);
    system(systemcall);
```

```
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_global_head2.txt >
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable3/%s.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2",home,group_string,home,gr
oup_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_tail2.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);

    if(debug) {
      sprintf(systemcall,"tail
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2",home,group_string);
      system(systemcall);
       }

    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.res ",home,home,group_string);
    system(systemcall);

        printf("\nFinished global WNL run\n");

    /* ORIGIN-SPECIFIC */

    /* Generate 'OS' flow file */
    printf("\nGenerating OS Traditional data for %s year old
migrants\n",group_string);
    sprintf(filename,"%sData/FlowData/SMSTable3/%s.txt",home,group_string);
    if ((in_flows=fopen(filename,"r"))==NULL)
    {
      return(printf("Error opening input flow file\n"));
    }
    sprintf(filename,"%sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    if ((out_flows=fopen(filename,"w"))==NULL)
    {
      return(printf("Error opening output flow file\n"));
    }
    while(fgets(inString,100,in_flows))
    {
      sscanf(inString,"%10i%10i%*5i%5i",&interaction,&separation,&destination);
      fprintf(out_flows,"%10i%10i    1%5i\n",interaction,separation,destination);
    }
    fclose(in_flows);
    fclose(out_flows);

    /* MAIN ORIGIN PROGRESSION LOOP */
    for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
    {
        // Determine which district correlates with loop_num which is an nl#
      for(loop1=0;loop1<460;loop1++)
              if(district_nl_index[loop1]==loop_num)
                origin_dist=loop1;
         printf("\nOrigin_dist# set to %i\n",origin_dist);

        // Stage 1 calibration
      printf("Doing origin nl#%i stage1 run for %s\n",loop_num,group_string);
      sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_head1.txt >
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_n
um);
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable3/%s.os.txt >
%sData/FlowData/SMSTable3/os.tmp", (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable3/os.tmp >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run1",
home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_tail.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_n
um);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
```

```
     sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num,home);
     system(systemcall);
     sprintf(systemcall, "%sModels/SIModel/simodel",home);
     system(systemcall);
     sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.res1",home,home,group_string,loop_n
um);
     system(systemcall);

        // Open stage one results file and extract parameter estimates
     sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
     in_var = fopen(filename,"r");
     dist_pe = pop_pe = socclass_pe = hp_pe = tenure_pe = unemp_pe = 99999.0;
     for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
       fgets(inString, 100, in_var);
     fgets(inString, 100, in_var);
     sscanf(inString,"%*20c%121f", &dist_pe);
     fgets(inString, 100, in_var);
     sscanf(inString,"%*20c%121f", &pop_pe);
     fgets(inString, 100, in_var);
     sscanf(inString,"%*20c%121f", &socclass_pe);
     fgets(inString, 100, in_var);
     sscanf(inString,"%*20c%121f", &hp_pe);
     fgets(inString, 100, in_var);
     sscanf(inString,"%*20c%121f", &tenure_pe);
     fgets(inString, 100, in_var);
     sscanf(inString,"%*20c%121f", &unemp_pe);
        fclose(in_var);
     sscanf(inString,"%*20c%121f", &access_pe);
        fclose(in_var);

     // Read appropriate Weighted regionalisation into regionalization[][] variable
     sprintf(filename,"%sProgs/Regionaliser/%i.matrix",home,loop_num);
     in_reg = fopen(filename,"r");
     for(loop2=1;loop2<460;loop2++)
          for(loop3=1;loop3<460;loop3++)
        fscanf(in_reg,"%i%*c",&regionalization[loop2][loop3]);
     fclose(in_reg);
     if(debug)
        {
          printf("Weighted Regionalization Matrix:-\n");
       for(loop2=1;loop2<5;loop2++)
             {
            for(loop3=1;loop3<6;loop3++)
          printf("%i,",regionalization[loop2][loop3]);
        printf("%i\n",regionalization[loop2][5]);
             }
             printf("\n");
        }

     // Calculate utility of each individual district
     utility[origin_dist] = 0.0;
     for(loop2=1;loop2<460;loop2++)
             if(loop2!=origin_dist)
               utility[loop2] = (    dist_pe *
log((double)sepmat[origin_dist][loop2]) ) +   (pop_pe * log((double)pop[loop2]) ) +
                                  (socclass_pe * log(socclass[loop2])              ) +
(hp_pe * log((double)hp[loop2])  ) +
                                  (  tenure_pe * log(tenure[loop2])                ) +
(unemp_pe * log(unemp[loop2])        ) +
                                                   (  access_pe * log(access[loop2])
);

      if(debug && loop_num<5)
        for(loop2=1;loop2<11;loop2++)
          for(loop3=1;loop3<460;loop3++)
            if(district_nl_index[loop3]==loop2) {
              printf("ParaEstimates   : %f\t%f\t%f\t%f\t%f\t%f\t%f\n",dist_pe,pop_pe,
socclass_pe, hp_pe, tenure_pe, unemp_pe, access_pe);
              printf("Variables[nl#%i]:
%f\t%f\t%f\t%f\t%f\t%f\t%f\n",loop2,(double)sepmat[origin_dist][loop3],(double)pop[loo
p3],socclass[loop3],(double)hp[loop3],tenure[loop3],unemp[loop3], access[loop3]);
                 printf("utility[nl#%i]   : %f\n\n",loop2,utility[loop3]);
                    }

      // Calculate inclusive_value[]'s  -  summing weighted exp(utility) of ALL
districts
      for(loop2=1;loop2<460;loop2++)
        rutility[loop2]=0.0;
```

```
      for(loop2=1;loop2<460;loop2++)
        for(loop3=1;loop3<460;loop3++)

rutility[loop2]+=(utility[loop3]*(double)(regionalization[loop2][loop3])/1000.0);
      if(debug && loop_num<5)
        for(loop2=1;loop2<11;loop2++)
          for(loop3=1;loop3<460;loop3++)
            if(district_nl_index[loop3]==loop2)
                 printf("rutility[nl%i] = %f\n",loop2,rutility[loop3]);

      // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
      sum = mean = range = 0.0;
         min = max = rutility[1];
      for(loop2=1;loop2<460;loop2++)
         {
           sum += rutility[loop2];
               if(rutility[loop2]<min) min = rutility[loop2];
               if(rutility[loop2]>max) max = rutility[loop2];
         }
         mean = sum / 459.0;
         range = max - min;
         printf("Mean of regional utility = %f\n",mean);
         printf("Range of regional utility = %f\n",range);
      if(abs(mean)>range)
           for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/abs(mean);
         else
           for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/range;
      if(debug && loop_num<5)
        for(loop2=1;loop2<11;loop2++)
          for(loop3=1;loop3<460;loop3++)
            if(district_nl_index[loop3]==loop2)
                 printf("normalized rutility[nl%i] = %f\n",loop2,rutility[loop3]);

         // exp() rutility[]
         for(loop2=1;loop2<460;loop2++) {
       rutility[loop2]=exp(rutility[loop2]);
              if(abs(rutility[loop2])>1000.0) {
                printf("rutility[%i]=%f\nThis is a suspiciously high regional utility
value, outputing diagnostics:\n",loop2,rutility[loop2]);
                diagnostics = 1;
                pause = 1;
              }
         }

         if(diagnostics) {
           for(loop3=1;loop3<460;loop3++) {
          printf("OS DNL FOR nl#%i for %s:\n",loop_num,group_string);
               printf("PARAMS:
%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\n",dist_pe,pop_pe,socclass_pe,hp_pe,te
nure_pe,unemp_pe,access_pe);
                 printf("VALUES:
%i\t%i\t%5.2f\t%d\t%5.2f\t%5.2f\t%5.2f\n",sepmat[1][loop3],pop[loop3],socclass[loop3],
hp[loop3],tenure[loop3],unemp[loop3],access[loop3]);
                 printf("district %i utility = %8.3f + %8.3f + %8.3f + %8.3f + %8.3f +
%8.3f + %8.3f = %8.3f\n", loop3,
                                                   (dist_pe     *
log((double)sepmat[origin_dist][loop3]),(pop_pe   * log((double)pop[loop3])),
                                   (socclass_pe * log(socclass[loop3])),
(hp_pe * log((double)hp[loop3])),
                                   (tenure_pe   * log(tenure[loop3])),
(unemp_pe * log(unemp[loop3])) ,
                                                   (access_pe   * log(access[loop3])),
utility[loop3] );
              }
        printf("\n");
         }

       if(debug && loop_num<5)
         for(loop2=1;loop2<11;loop2++)
           for(loop3=1;loop3<460;loop3++)
             if(district_nl_index[loop3]==loop2)
                  printf("exp'd normalized rutility[nl%i] =
%f\n",loop2,rutility[loop3]);

       // Output inclusive values for selected 100 districts (in NL order) to file
       sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
       system(systemcall);
       sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
```

```
    if((out_rutilities = fopen(filename,"w"))==NULL)
        return(printf("Couldn't open %s for writing, exiting...\n",filename));
    for(loop2=1;loop2<101;loop2++)
      for(loop3=1;loop3<460;loop3++)
        if( district_nl_index[loop3] == loop2)
          fprintf(out_rutilities,"%f\n",rutility[loop3]);
    fclose(out_rutilities);

    // SECOND stage calibration
    printf("\nDoing origin nl#%i stage2 run for %s\n",loop_num,group_string);

    sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/wnl_tail.txt
%sModels/HybridWeightedNestedLogit/wnl_tail2.txt",home,home);
    system(systemcall);
    sprintf(systemcall,"cat rutilities.csv >>
%sModels/HybridWeightedNestedLogit/wnl_tail2.txt",home);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_head2.txt >
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_n
um);
    system(systemcall);
    sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable3/%s.os.txt >
%sData/FlowData/SMSTable3/os.tmp", (loop_num-1)*100+1,home,group_string,home);
    system(systemcall);
    sprintf(systemcall,"head -100 %sData/FlowData/SMSTable3/os.tmp >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run2",
home,home,group_string,loop_num);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_tail2.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_n
um);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num, home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.res
",home,home,group_string,loop_num);
    system(systemcall);

    printf("\nFinished origin nl#%i WNL run\n", loop_num);

        if((debug)&&(loop_num<4)) {
                printf("In debug stepping mode... press any key to
continue...\n",loop_num,group_string);
                getchar();
        }

        if(pause) {
          printf("Pausing due to output of diagnostic info, press any key to
continue...\n");
          getchar();
          pause = 0;
              diagnostics = 0;
        }

    } /* END OF ORIGIN PROGRESSION LOOP */

    sprintf(systemcall,"rm %sData/FlowData/SMSTable3/%s.os.txt",home,group_string);
    system(systemcall);
       printf("Completed all calibrations for group: %s\n\n",group_string);

    if(debug)
          return(printf("Terminating after first set of global + OS runs because
running in debug mode\n"));

  } /* END OF sub_group PROGRESSION LOOP */

  // AGE-GROUP PROGRESSION LOOP
  for(sub_group=0;sub_group<9;sub_group++)
  {
    if (sub_group == 0)
      strcpy(group_string,"all_single");
    else if (sub_group == 1)
      strcpy(group_string,"all_married");
    else if (sub_group == 2)
      strcpy(group_string,"all_wid_div");
```

```
    else if (sub_group == 3)
      strcpy(group_string,"male_single");
    else if (sub_group == 4)
      strcpy(group_string,"male_married");
    else if (sub_group == 5)
      strcpy(group_string,"male_wid_div");
    else if (sub_group == 6)
      strcpy(group_string,"female_single");
    else if (sub_group == 7)
      strcpy(group_string,"female_married");
    else
      strcpy(group_string,"female_wid_div");

        // Create sub-group output directory within LatestRun directory
        sprintf(systemcall,"mkdir
%sModels/HybridWeightedNestedLogit/LatestRun/%s",home,group_string);
    system(systemcall);

        // GLOBAL CALIBRATION - using best Weighted regionalization calculated for
origin #1 (arbitrary decision - global WNL does really make sense)

    // Set origin_dist to dist# for nl#1
        for (loop2=1;loop2<460;loop2++)
      if(district_nl_index[loop2] == 1) {
        origin_dist = loop2;
        break;
          }
        printf("Origin_dist# set to %i\n",origin_dist);

        // Stage 1 calibration
    printf("\nDoing global stage1 run for %s\n",group_string);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_global_head1.txt >
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable4/%s.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run1",home,group_string,home,gr
oup_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_tail.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run1",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.res1 ",home,home,group_string);
    system(systemcall);

        // Open stage one results file and extract parameter estimates
    sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
    in_var = fopen(filename,"r");
    dist_pe = pop_pe = socclass_pe = hp_pe = tenure_pe = unemp_pe = 99999.0;
    for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
      fgets(inString, 100, in_var);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%121f", &dist_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%121f", &pop_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%121f", &socclass_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%121f", &hp_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%121f", &tenure_pe);
    fgets(inString, 100, in_var);
    sscanf(inString,"%*20c%121f", &unemp_pe);
    fclose(in_var);
    sscanf(inString,"%*20c%121f", &access_pe);
    fclose(in_var);

    // Read appropriate Weighted regionalisation into regionalization[][] variable
    sprintf(filename,"%sProgs/Regionaliser/1.matrix",home);
    in_reg = fopen(filename,"r");
    for(loop2=1;loop2<460;loop2++)
          for(loop3=1;loop3<460;loop3++)
        fscanf(in_reg,"%i%*c",&regionalization[loop2][loop3]);
```

```
    fclose(in_reg);
    if(debug)
        {
          printf("Weighted Regionalization Matrix:-\n");
      for(loop2=1;loop2<5;loop2++)
          {
            for(loop3=1;loop3<5;loop3++)
          printf("%i,",regionalization[loop2][loop3]);
        printf("%i\n",regionalization[loop2][5]);
          }
        }

    // Calculate utility of each individual district
    utility[origin_dist] = 0.0;
    for(loop2=1;loop2<460;loop2++)
          if (loop2 != origin_dist) {
            utility[loop2] = (     dist_pe * log((double)sepmat[origin_dist][loop2]) ) +
(pop_pe * log((double)pop[loop2]) ) +
                                   (socclass_pe * log(socclass[loop2])           ) +    (hp_pe
* log((double)hp[loop2])  ) +
                                   (  tenure_pe * log(tenure[loop2])             ) + (unemp_pe
* log(unemp[loop2])       ) +
                                                    (  access_pe * log(access[loop2])
);
          }

    if (debug)
          for(loop2=1;loop2<460;loop2+=25)
            printf("utility[%i] = %f\n",loop2,utility[loop2]);

    // Calculate inclusive_value[]'s  -  summing weighted exp(utility) of ALL
districts
    for(loop2=1;loop2<460;loop2++)
      rutility[loop2]=0.0;
    for(loop2=1;loop2<460;loop2++)
      for(loop3=1;loop3<460;loop3++)

rutility[loop2]+=(utility[loop3]*(double)(regionalization[loop2][loop3])/1000.0);
        if(debug)
      for(loop2=1;loop2<5;loop2++)
            printf("rutility[%i] = %f\n",loop2,rutility[loop2]);

    // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
    sum = mean = range = 0.0;
    min = max = rutility[1];
    for(loop2=1;loop2<460;loop2++)
        {
          sum += rutility[loop2];
          if(rutility[loop2]<min) min = rutility[loop2];
      if(rutility[loop2]>max) max = rutility[loop2];
        }
        mean = sum / 459.0;
        range = max - min;
        printf("Mean of regional utility = %f\n",mean);
        printf("Range of regional utility = %f\n",range);
    if(abs(mean)>range)
          for(loop2=1;loop2<460;loop2++)
         rutility[loop2]=rutility[loop2]/abs(mean);
        else
          for(loop2=1;loop2<460;loop2++)
         rutility[loop2]=rutility[loop2]/range;

        // Apply exp() to rutility[]
        for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=exp(rutility[loop2]);

     /* Output inclusive values for selected 100 districts (in NL order) to file */
     sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
     system(systemcall);
     sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
     if((out_rutilities = fopen(filename,"w"))==NULL)
           return(printf("Couldn't open %s for writing, exiting...\n",filename));
     for(loop2=1;loop2<101;loop2++)
       for(loop3=1;loop3<460;loop3++)
         if( district_nl_index[loop3] == loop2)
           fprintf(out_rutilities,"%f\n",rutility[loop3]);
        fclose(out_rutilities);

     // SECOND stage calibration
     printf("\nDoing global stage2 run for %s\n",group_string);
```

```
    sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/wnl_tail.txt
%sModels/HybridWeightedNestedLogit/wnl_tail2.txt",home,home);
    system(systemcall);
    sprintf(systemcall,"cat rutilities.csv >>
%sModels/HybridWeightedNestedLogit/wnl_tail2.txt",home);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_global_head2.txt >
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"cat %sData/FlowData/SMSTable4/%s.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2",home,group_string,home,gr
oup_string);
    system(systemcall);
    sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_tail2.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2",home,home,group_string);
    system(systemcall);
    sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
    system(systemcall);
    sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/LatestRun/%s/global.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,home);
    system(systemcall);
    sprintf(systemcall, "%sModels/SIModel/simodel",home);
    system(systemcall);
    sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/HybridWeightedNestedLogit/LatestRun/%s/global.res ",home,home,group_string);
    system(systemcall);

        printf("\nFinished global WNL run\n");

    /* ORIGIN-SPECIFIC */

    /* Generate 'OS' flow file */
    printf("\nGenerating OS Traditional data for %s year old
migrants\n",group_string);
    sprintf(filename,"%sData/FlowData/SMSTable4/%s.txt",home,group_string);
    if ((in_flows=fopen(filename,"r"))==NULL)
    {
      return(printf("Error opening input flow file\n"));
    }
    sprintf(filename,"%sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
    if ((out_flows=fopen(filename,"w"))==NULL)
    {
      return(printf("Error opening output flow file\n"));
    }
    while(fgets(inString,100,in_flows))
    {
      sscanf(inString,"%10i%10i%*5i%5i",&interaction,&separation,&destination);
      fprintf(out_flows,"%10i%10i    1%5i\n",interaction,separation,destination);
    }
    fclose(in_flows);
    fclose(out_flows);

    /* MAIN ORIGIN PROGRESSION LOOP */
    for(loop_num=START_AREA;loop_num<=STOP_AREA;loop_num+=STEP_BY)
    {
        // Determine which district correlates with loop_num which is an nl#
      for(loop1=0;loop1<460;loop1++)
              if(district_nl_index[loop1]==loop_num)
                origin_dist=loop1;
        printf("\nOrigin_dist# set to %i\n",origin_dist);

        // Stage 1 calibration
      printf("Doing origin nl#%i stage1 run for %s\n",loop_num,group_string);
      sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_head1.txt >
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_n
um);
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable4/%s.os.txt >
%sData/FlowData/SMSTable4/os.tmp",(loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable4/os.tmp >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_n
um);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_tail.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run1",home,home,group_string,loop_n
um);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
```

```
      sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run1
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num,home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/SIModel/simodel2_0.tst
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.res1",home,home,group_string,loop_n
um);
      system(systemcall);

         // Open stage one results file and extract parameter estimates
      sprintf(filename,"%sModels/SIModel/simodel2_0.tst",home);
      in_var = fopen(filename,"r");
      dist_pe = pop_pe = socclass_pe = hp_pe = tenure_pe = unemp_pe = 99999.0;
      for(loop2=0;loop2<34;loop2++) // Skip 34 preliminary lines in the SIModel output
file
        fgets(inString, 100, in_var);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%121f", &dist_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%121f", &pop_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%121f", &socclass_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%121f", &hp_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%121f", &tenure_pe);
      fgets(inString, 100, in_var);
      sscanf(inString,"%*20c%121f", &unemp_pe);
         fclose(in_var);
      sscanf(inString,"%*20c%121f", &access_pe);
         fclose(in_var);

     // Read appropriate Weighted regionalisation into regionalization[][] variable
      sprintf(filename,"%sProgs/Regionaliser/%i.matrix",home,loop_num);
      in_reg = fopen(filename,"r");
      for(loop2=1;loop2<460;loop2++)
          for(loop3=1;loop3<460;loop3++)
         fscanf(in_reg,"%i%*c",&regionalization[loop2][loop3]);
      fclose(in_reg);
      if(debug)
         {
           printf("Weighted Regionalization Matrix:-\n");
        for(loop2=1;loop2<5;loop2++)
             {
              for(loop3=1;loop3<5;loop3++)
            printf("%i,",regionalization[loop2][loop3]);
          printf("%i\n",regionalization[loop2][5]);
             }
         }

      // Calculate utility of each individual district
      utility[origin_dist] = 0.0;
      for(loop2=1;loop2<460;loop2++)
        if(loop2!=origin_dist)
                 utility[loop2] = (    dist_pe *
log((double)sepmat[origin_dist][loop2]) ) +   (pop_pe * log((double)pop[loop2]) ) +
                              (socclass_pe * log(socclass[loop2])               ) +
(hp_pe * log((double)hp[loop2])  ) +
                              (  tenure_pe * log(tenure[loop2])                 ) +
(unemp_pe * log(unemp[loop2])        ) +
                                        (  access_pe * log(access[loop2])
);
      if (debug)
           for(loop2=1;loop2<460;loop2+=25)
             printf("utility[%i] = %f\n",loop2,utility[loop2]);

      // Calculate inclusive_value[]'s  -  summing weighted exp(utility) of ALL
districts
      for(loop2=1;loop2<460;loop2++)
        rutility[loop2]=0.0;
      for(loop2=1;loop2<460;loop2++)
        for(loop3=1;loop3<460;loop3++)

rutility[loop2]+=(utility[loop3]*(double)(regionalization[loop2][loop3])/1000.0);

      if(debug)
        for(loop2=1;loop2<5;loop2++)
             printf("rutility[%i] = %f\n",loop2,rutility[loop2]);
```

```
    // Normalise rutility[] so that exp(rutility[]) calls generate smaller range of
values.
      sum = mean = range = 0.0;
         min = max = rutility[1];
      for(loop2=1;loop2<460;loop2++)
         {
           sum += rutility[loop2];
               if(rutility[loop2]<min) min = rutility[loop2];
               if(rutility[loop2]>max) max = rutility[loop2];
         }
         mean = sum / 459.0;
         range = max - min;
         printf("Mean of regional utility = %f\n",mean);
         printf("Range of regional utility = %f\n",range);
      if(abs(mean)>range)
           for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/abs(mean);
         else
           for(loop2=1;loop2<460;loop2++)
          rutility[loop2]=rutility[loop2]/range;

                       // exp() rutility[]
         for(loop2=1;loop2<460;loop2++) {
        rutility[loop2]=exp(rutility[loop2]);
                if(abs(rutility[loop2])>1000.0) {
                  printf("rutility[%i]=%f\nThis is a suspiciously high regional utility
value, outputing diagnostics:\n",loop2,rutility[loop2]);
                  diagnostics = 1;
                  pause = 1;
                }
         }

         if(diagnostics) {
           for(loop3=1;loop3<460;loop3++) {
          printf("OS DNL FOR nl#%i for %s:\n",loop_num,group_string);
              printf("PARAMS:
%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\t%5.3f\n",dist_pe,pop_pe,socclass_pe,hp_pe,te
nure_pe,unemp_pe,access_pe);
                  printf("VALUES:
%i\t%i\t%5.2f\t%d\t%5.2f\t%5.2f\t%5.2f\n",sepmat[1][loop3],pop[loop3],socclass[loop3],
hp[loop3],tenure[loop3],unemp[loop3],access[loop3]);
                  printf("district %i utility = %8.3f + %8.3f + %8.3f + %8.3f + %8.3f +
%8.3f + %8.3f = %8.3f\n", loop3,
                                               (dist_pe     *
log((double)sepmat[origin_dist][loop3])),(pop_pe    * log((double)pop[loop3])),
                              (socclass_pe * log(socclass[loop3])),
 (hp_pe * log((double)hp[loop3])),
                              (tenure_pe    * log(tenure[loop3])),
 (unemp_pe * log(unemp[loop3])) ,
                                               (access_pe    * log(access[loop3])),
utility[loop3]);
                 }
        printf("\n");
          }


          /* Output inclusive values for selected 100 districts (in NL order) to file
 */
       sprintf(systemcall,"rm %s/Models/SIModel/rutilities.csv",home);
       system(systemcall);
       sprintf(filename,"%s/Models/SIModel/rutilities.csv",home);
       if((out_rutilities = fopen(filename,"w"))==NULL)
           return(printf("Couldn't open %s for writing, exiting...\n",filename));
       for(loop2=1;loop2<101;loop2++)
         for(loop3=1;loop3<460;loop3++)
           if( district_nl_index[loop3] == loop2)
             fprintf(out_rutilities,"%f\n",rutility[loop3]);
          fclose(out_rutilities);

       // SECOND stage calibration
       printf("\nDoing origin nl#%i stage2 run for %s\n",loop_num,group_string);

       sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/wnl_tail.txt
 %sModels/HybridWeightedNestedLogit/wnl_tail2.txt",home,home);
       system(systemcall);
       sprintf(systemcall,"cat rutilities.csv >>
 %sModels/HybridWeightedNestedLogit/wnl_tail2.txt",home);
       system(systemcall);
       sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_head2.txt >
 %sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_n
 um);
```

```
      system(systemcall);
      sprintf(systemcall,"tail +%i %sData/FlowData/SMSTable4/%s.os.txt >
%sData/FlowData/SMSTable4/os.tmp", (loop_num-1)*100+1,home,group_string,home);
      system(systemcall);
      sprintf(systemcall,"head -100 %sData/FlowData/SMSTable4/os.tmp >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run2",
home,home,group_string,loop_num);
      system(systemcall);
      sprintf(systemcall,"cat %sModels/HybridWeightedNestedLogit/wnl_tail2.txt >>
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run2",home,home,group_string,loop_n
um);
      system(systemcall);
      sprintf(systemcall,"rm %sModels/SIModel/simodel2_0.*",home);
      system(systemcall);
      sprintf(systemcall,"cp %sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.run2
%sModels/SIModel/simodel2_0.dat",home,group_string,loop_num, home);
      system(systemcall);
      sprintf(systemcall, "%sModels/SIModel/simodel",home);
      system(systemcall);
      sprintf(systemcall,"mv %sModels/SIModel/simodel2_0.tst
%sModels/HybridWeightedNestedLogit/LatestRun/%s/%i.res
",home,home,group_string,loop_num);
      system(systemcall);

      printf("\nFinished origin nl#%i WNL run\n", loop_num);

          if((debug)&&(loop_num<4)) {
                  printf("In debug stepping mode... press any key to
continue...\n",loop_num,group_string);
                  getchar();
          }

          if(pause) {
            printf("Pausing due to output of diagnostic info, press any key to
continue...\n");
            getchar();
            pause = 0;
                diagnostics = 0;
          }

     } /* END OF ORIGIN PROGRESSION LOOP */

     sprintf(systemcall,"rm %sData/FlowData/SMSTable4/%s.os.txt",home,group_string);
     system(systemcall);
         printf("Completed all calibrations for group: %s\n\n",group_string);
   } /* END OF sub_group PROGRESSION LOOP */

   printf("Finished all WNL calibration runs!\n\n");

   return;
 }

 void initialiseNLAreas()
 {
   int loop;
   for(loop=1;loop<460;loop++)
     district_nl_index[loop]=0;

   district_nl_index[418]=1;
   district_nl_index[15]=2;
   district_nl_index[49]=3;
   district_nl_index[70]=4;
   district_nl_index[58]=5;
   district_nl_index[241]=6;
   district_nl_index[242]=7;
   district_nl_index[34]=8;
   district_nl_index[141]=9;
   district_nl_index[65]=10;
   district_nl_index[157]=11;
   district_nl_index[71]=12;
   district_nl_index[35]=13;
   district_nl_index[91]=14;
   district_nl_index[2]=15;
   district_nl_index[228]=16;
   district_nl_index[398]=17;
   district_nl_index[118]=18;
   district_nl_index[168]=19;
   district_nl_index[178]=20;
   district_nl_index[97]=21;
   district_nl_index[169]=22;
   district_nl_index[59]=23;
```

```
    district_nl_index[20]=24;
    district_nl_index[150]=25;
    district_nl_index[125]=26;
    district_nl_index[50]=27;
    district_nl_index[230]=28;
    district_nl_index[60]=29;
    district_nl_index[455]=30;
    district_nl_index[152]=31;
    district_nl_index[21]=32;
    district_nl_index[432]=33;
    district_nl_index[132]=34;
    district_nl_index[445]=35;
    district_nl_index[181]=36;
    district_nl_index[23]=37;
    district_nl_index[341]=38;
    district_nl_index[3]=39;
    district_nl_index[4]=40;
    district_nl_index[293]=41;
    district_nl_index[24]=42;
    district_nl_index[334]=43;
    district_nl_index[6]=44;
    district_nl_index[7]=45;
    district_nl_index[274]=46;
    district_nl_index[223]=47;
    district_nl_index[8]=48;
    district_nl_index[247]=49;
    district_nl_index[68]=50;
    district_nl_index[259]=51;
    district_nl_index[266]=52;
    district_nl_index[45]=53;
    district_nl_index[76]=54;
    district_nl_index[102]=55;
    district_nl_index[233]=56;
    district_nl_index[36]=57;
    district_nl_index[107]=58;
    district_nl_index[88]=59;
    district_nl_index[81]=60;
    district_nl_index[54]=61;
    district_nl_index[382]=62;
    district_nl_index[282]=63;
    district_nl_index[276]=64;
    district_nl_index[305]=65;
    district_nl_index[37]=66;
    district_nl_index[308]=67;
    district_nl_index[95]=68;
    district_nl_index[135]=69;
    district_nl_index[145]=70;
    district_nl_index[192]=71;
    district_nl_index[249]=72;
    district_nl_index[82]=73;
    district_nl_index[38]=74;
    district_nl_index[234]=75;
    district_nl_index[51]=76;
    district_nl_index[39]=77;
    district_nl_index[296]=78;
    district_nl_index[52]=79;
    district_nl_index[194]=80;
    district_nl_index[11]=81;
    district_nl_index[211]=82;
    district_nl_index[328]=83;
    district_nl_index[410]=84;
    district_nl_index[40]=85;
    district_nl_index[330]=86;
    district_nl_index[353]=87;
    district_nl_index[57]=88;
    district_nl_index[403]=89;
    district_nl_index[365]=90;
    district_nl_index[12]=91;
    district_nl_index[42]=92;
    district_nl_index[69]=93;
    district_nl_index[63]=94;
    district_nl_index[104]=95;
    district_nl_index[354]=96;
    district_nl_index[43]=97;
    district_nl_index[85]=98;
    district_nl_index[64]=99;
    district_nl_index[298]=100;
}
```

Appendix D: Traditional model goodness-of-fit statistics and parameter estimates values

This appendix tabulates in full the AIC and R^2_{adj} goodness-of-fit statistics and the parameter estimates (with their standard errors) arising from global and origin-specific calibrations of the traditional migration destination choice model for all migrants aged 16+ years.

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	R2adj	AIC
Global	-1.24	0.002	0.78	0.005	0.92	0.019	-0.50	0.013	-0.27	0.016	-0.27	0.015	0.810	92587.8
Aberdeen City	-2.19	0.099	1.15	0.059	3.07	0.328	0.64	0.254	0.12	0.307	0.95	0.253	0.939	451.6
Barking and Dagenham	-1.21	0.073	0.39	0.126	0.34	0.364	-2.22	0.329	-1.04	0.249	-1.48	0.264	0.473	389.7
Barnsley	-1.83	0.049	0.38	0.069	-0.84	0.292	1.03	0.270	0.71	0.263	0.60	0.197	0.931	515.1
Bath	-1.13	0.036	0.95	0.077	1.38	0.275	0.21	0.187	-0.83	0.237	0.00	0.210	0.917	363.4
Birmingham	-1.24	0.015	0.99	0.039	0.70	0.127	0.05	0.086	-1.09	0.105	-0.44	0.092	0.881	750.5
Blackburn	-1.11	0.042	0.46	0.089	-1.29	0.348	2.00	0.250	1.72	0.334	1.16	0.279	0.636	451.7
Blackpool	-1.20	0.042	0.45	0.064	0.15	0.265	0.04	0.171	0.27	0.272	0.21	0.210	0.728	447.0
Bolton	-1.39	0.021	0.43	0.062	-1.57	0.224	0.87	0.173	-0.12	0.217	-0.55	0.181	0.909	609.1
Bournemouth	-0.81	0.033	0.86	0.077	0.28	0.254	0.41	0.187	-0.35	0.226	-0.24	0.201	0.794	367.6
Bradford	-1.11	0.029	1.39	0.039	-0.51	0.183	0.64	0.128	-1.69	0.150	-1.47	0.124	0.946	669.4
Brighton	-0.93	0.083	0.95	0.068	1.79	0.243	-0.54	0.230	-0.51	0.181	0.45	0.180	0.745	398.8
Bristol	-0.93	0.031	0.81	0.046	0.94	0.152	0.45	0.103	-0.62	0.142	0.22	0.126	0.668	551.3
Bury	-1.51	0.024	0.24	0.071	1.75	0.278	-0.32	0.164	-1.28	0.233	0.74	0.242	0.965	525.8
Cambridge	-0.25	0.067	0.74	0.058	1.71	0.210	0.39	0.161	-0.79	0.153	0.51	0.150	0.588	498.9
Camden	-0.83	0.027	0.71	0.052	2.57	0.133	-1.16	0.121	-0.77	0.118	0.56	0.111	0.971	587.8
Canterbury	-1.69	0.049	0.39	0.084	1.71	0.243	-1.63	0.192	0.70	0.205	0.97	0.206	0.873	476.3
Cardiff	-1.22	0.028	0.68	0.049	2.47	0.205	-0.46	0.118	0.04	0.194	1.11	0.167	0.891	475.8
Carlisle	-1.44	0.130	0.56	0.083	1.10	0.366	-0.04	0.266	-0.58	0.308	-0.12	0.257	0.567	312.7
Chelmsford	-1.64	0.071	0.65	0.092	1.50	0.248	-2.46	0.233	-0.66	0.211	-1.11	0.208	0.521	513.7
Cheltenham	-1.69	0.042	0.54	0.078	1.78	0.341	-0.81	0.224	-0.71	0.294	0.28	0.246	0.949	360.1
Chester	-0.80	0.047	0.83	0.068	1.89	0.260	-0.08	0.158	-1.12	0.218	0.23	0.195	0.740	376.3
Colchester	-1.21	0.052	0.62	0.077	1.58	0.235	-1.51	0.173	-0.93	0.195	-0.55	0.188	0.697	437.2
Coventry	-1.39	0.023	0.36	0.044	2.74	0.204	-0.57	0.130	0.30	0.180	1.15	0.157	0.910	596.1
Croydon	-1.24	0.037	0.82	0.057	2.49	0.152	-2.07	0.155	0.30	0.127	0.69	0.129	0.804	693.3
Darlington	-1.24	0.074	0.63	0.086	0.87	0.380	0.23	0.275	-0.84	0.313	-0.20	0.241	0.767	268.8
Derby	-0.98	0.056	0.61	0.069	0.64	0.261	0.07	0.166	-0.23	0.242	0.53	0.218	0.726	403.5

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	R2adj	AIC
Doncaster	-1.65	0.039	0.35	0.053	0.08	0.229	0.25	0.165	-0.43	0.226	0.06	0.179	0.891	550.0
Dover	-1.66	0.046	0.58	0.115	1.22	0.308	-1.35	0.227	0.67	0.299	0.34	0.273	0.944	366.9
Dudley	-1.45	0.031	0.42	0.050	0.51	0.306	0.09	0.211	0.78	0.300	1.22	0.240	0.950	509.7
Dundee City	-1.69	0.086	1.13	0.069	0.98	0.412	0.62	0.290	-1.84	0.181	-0.76	0.131	0.975	331.4
Durham	-1.32	0.050	0.75	0.071	1.39	0.258	0.69	0.191	-1.16	0.244	0.05	0.198	0.924	331.1
Ealing	-1.10	0.028	0.62	0.049	2.70	0.148	-1.48	0.140	1.03	0.121	1.38	0.121	0.836	698.6
Edinburgh City	-1.01	0.055	0.89	0.048	2.56	0.204	-0.56	0.144	-2.84	0.098	-0.92	0.084	0.840	551.0
Exeter	-1.56	0.062	0.83	0.078	0.28	0.229	0.61	0.179	-0.83	0.240	-0.21	0.208	0.689	422.6
Glasgow City	-1.03	0.131	1.24	0.120	1.59	0.683	-0.34	0.436	-2.10	0.347	-1.21	0.306	0.789	224.5
Gloucester	-1.74	0.042	0.61	0.091	0.65	0.366	-0.50	0.259	-0.75	0.352	-0.09	0.288	0.964	331.0
Greenwich	-0.71	0.039	0.61	0.066	0.59	0.181	0.11	0.189	0.39	0.155	0.54	0.153	0.653	564.9
Guildford	-0.41	0.068	0.68	0.076	0.91	0.247	0.69	0.238	-0.12	0.193	0.27	0.187	0.623	385.3
Hackney	-0.81	0.026	0.70	0.063	0.91	0.126	-0.54	0.111	-1.00	0.149	0.10	0.144	0.979	543.2
Hammersmith and Fulham	-1.01	0.031	0.95	0.055	1.65	0.150	-0.18	0.164	1.15	0.122	1.29	0.143	0.892	766.7
Harrogate	-1.65	0.042	0.57	0.046	-0.16	0.238	0.58	0.164	-1.76	0.197	-1.38	0.168	0.924	526.9
Harrow	-1.15	0.040	0.77	0.070	1.61	0.207	-0.99	0.202	0.95	0.161	0.68	0.166	0.894	532.0
Ipswich	-1.55	0.076	0.68	0.122	1.71	0.382	-1.72	0.265	-1.39	0.301	-0.59	0.301	0.817	234.0
Islington	-0.85	0.030	0.70	0.058	2.08	0.163	-0.95	0.163	-0.73	0.131	0.52	0.128	0.952	642.1
Kensington and Chelsea	-1.11	0.025	0.29	0.060	2.89	0.171	-1.20	0.141	0.92	0.120	1.65	0.147	0.963	696.5
Kings Lynn & W. Norfolk	-2.30	0.116	0.30	0.163	-1.13	0.382	-0.20	0.312	-2.07	0.360	-1.83	0.358	0.435	441.9
Kingston upon Hull	-1.20	0.071	0.90	0.056	0.92	0.254	0.14	0.162	-0.65	0.210	0.08	0.186	0.774	385.6
Lambeth	-1.26	0.019	1.29	0.043	2.36	0.132	-1.86	0.124	0.01	0.093	0.14	0.102	0.955	738.7
Lancaster	-1.15	0.045	0.67	0.057	0.32	0.235	0.67	0.148	-0.33	0.212	0.05	0.177	0.748	416.0
Leeds	-1.45	0.017	1.08	0.035	0.01	0.124	0.74	0.083	-1.67	0.091	-1.05	0.079	0.914	820.5
Leicester	-0.72	0.045	0.91	0.048	1.44	0.203	0.16	0.122	0.23	0.167	0.95	0.155	0.675	492.3
Lincoln	-1.10	0.099	0.74	0.090	0.67	0.344	-0.13	0.218	0.03	0.318	0.13	0.276	0.526	341.4
Liverpool	-0.85	0.027	0.84	0.044	1.21	0.177	0.17	0.111	-0.82	0.154	0.22	0.142	0.711	551.5
Luton	-1.66	0.048	0.92	0.081	1.66	0.268	-2.13	0.185	-0.13	0.227	-0.38	0.216	0.913	406.8
Macclesfield	-1.39	0.031	0.72	0.069	2.63	0.290	-0.33	0.182	-2.01	0.230	-0.78	0.204	0.960	488.6
Maidstone	-1.57	0.055	0.08	0.098	0.21	0.226	-0.94	0.252	1.75	0.234	0.55	0.220	0.923	476.4
Manchester	-1.26	0.010	0.53	0.031	2.03	0.137	0.40	0.097	-0.11	0.119	1.09	0.101	0.949	817.2
Middlesbrough	-1.10	0.069	0.79	0.067	1.45	0.290	-0.21	0.206	-1.10	0.261	-0.06	0.217	0.735	338.4
Milton Keynes	-1.43	0.050	0.63	0.072	0.96	0.254	-1.14	0.181	-0.25	0.206	-0.22	0.194	0.776	441.0
Newbury	-2.49	0.048	0.36	0.089	0.79	0.258	-1.69	0.218	-0.40	0.229	-0.21	0.217	0.885	661.5

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	R2adj	AIC
Newcastle upon Tyne	-0.94	0.031	0.97	0.041	1.18	0.180	0.29	0.124	-1.21	0.137	-0.22	0.121	0.875	456.0
Newport	-1.47	0.050	0.67	0.103	1.91	0.379	-0.64	0.257	-0.45	0.388	0.45	0.312	0.948	336.2
Northampton	-1.48	0.047	0.61	0.069	1.32	0.262	-1.10	0.168	-0.12	0.226	0.09	0.207	0.811	428.5
Norwich	-2.84	0.117	0.80	0.143	1.02	0.456	-1.49	0.311	-0.98	0.411	-1.08	0.386	0.738	403.0
Nottingham	-1.04	0.040	0.77	0.049	1.44	0.197	0.17	0.120	-0.33	0.171	0.60	0.157	0.803	471.9
Oldham	-1.54	0.027	0.01	0.077	1.50	0.323	-0.15	0.197	-0.47	0.263	1.51	0.270	0.929	585.5
Oxford	-0.66	0.053	0.77	0.052	2.77	0.196	-0.39	0.144	-0.83	0.139	0.82	0.136	0.700	519.7
Peterborough	-1.34	0.084	0.56	0.085	0.08	0.284	-0.39	0.196	-0.71	0.252	-0.54	0.234	0.593	320.7
Plymouth	-1.43	0.051	0.81	0.060	-0.44	0.184	0.60	0.137	-0.29	0.186	-0.45	0.157	0.661	480.4
Poole	-1.25	0.051	0.80	0.146	1.10	0.469	-1.42	0.317	-1.33	0.438	-1.14	0.360	0.970	158.4
Portsmouth	-1.03	0.050	0.80	0.068	-0.20	0.214	-0.25	0.179	-0.12	0.193	-0.29	0.172	0.510	478.8
Preston	-1.17	0.041	0.38	0.073	-0.07	0.286	0.67	0.183	-0.12	0.270	0.42	0.231	0.735	442.3
Reading	-1.46	0.026	1.10	0.066	0.01	0.178	-1.59	0.144	-2.31	0.140	-2.66	0.121	0.916	754.1
Rochdale	-1.53	0.026	0.26	0.073	0.40	0.285	-0.31	0.189	-1.58	0.248	-0.23	0.236	0.916	578.3
Rochester upon Medway	-1.91	0.047	0.37	0.100	1.36	0.228	-2.51	0.214	0.87	0.245	0.21	0.230	0.883	525.6
Rotherham	-2.03	0.042	0.34	0.068	-0.67	0.314	0.93	0.226	1.14	0.259	1.11	0.220	0.983	517.3
Salford	-1.49	0.021	0.71	0.059	-1.53	0.227	0.89	0.172	-0.78	0.179	-0.60	0.170	0.883	730.0
Scarborough	-2.13	0.076	0.97	0.059	-0.77	0.270	0.81	0.213	-0.69	0.294	-1.15	0.202	0.891	355.0
Sheffield	-1.51	0.019	0.45	0.035	0.74	0.145	0.40	0.094	-0.49	0.127	0.77	0.114	0.887	776.5
Southampton	-1.29	0.052	0.62	0.072	1.55	0.234	-0.97	0.186	-0.11	0.204	0.48	0.174	0.798	387.3
Southwark	-1.05	0.021	0.93	0.055	1.31	0.129	-0.75	0.128	0.61	0.124	0.68	0.126	0.955	646.4
St Albans	-1.60	0.044	0.04	0.086	0.97	0.219	-2.11	0.197	1.35	0.197	0.70	0.200	0.751	580.5
Stafford	-1.28	0.051	0.91	0.074	-2.78	0.227	0.71	0.202	-1.52	0.271	-2.33	0.185	0.880	511.9
Stirling	-1.43	0.092	1.23	0.099	2.93	0.564	-0.45	0.362	-1.37	0.297	-0.77	0.241	0.966	261.3
Stockport	-1.40	0.020	0.80	0.056	2.91	0.245	-0.48	0.149	-1.95	0.170	-0.23	0.167	0.979	596.8
Stoke-on-Trent	-1.64	0.051	0.21	0.082	0.57	0.335	-1.11	0.189	-2.36	0.263	-1.46	0.242	0.759	579.8
Stratford-on-Avon	-1.91	0.037	0.56	0.074	2.45	0.368	-0.98	0.246	-1.14	0.311	-0.19	0.265	0.958	450.7
Sunderland	-1.20	0.047	0.82	0.057	0.42	0.262	0.45	0.184	-0.81	0.204	-0.24	0.172	0.935	386.6
Swansea	-1.68	0.071	0.85	0.076	1.72	0.288	0.24	0.194	0.06	0.294	0.63	0.243	0.909	318.5
Thamesdown	-1.53	0.064	0.63	0.074	0.52	0.276	-0.69	0.200	-0.50	0.236	-0.53	0.193	0.668	432.4
Tower Hamlets	-0.73	0.037	0.49	0.072	-0.92	0.158	0.72	0.179	-0.36	0.165	0.20	0.157	0.911	552.0
Trafford	-1.38	0.017	0.51	0.058	2.85	0.220	-0.79	0.126	-2.03	0.182	0.11	0.187	0.982	582.1
Wakefield	-1.67	0.042	0.84	0.048	-0.36	0.250	0.09	0.194	-1.05	0.212	-0.97	0.173	0.929	645.7
Walsall	-1.69	0.031	0.71	0.043	1.06	0.361	-0.11	0.228	0.10	0.316	1.20	0.249	0.985	485.9

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	R2adj	AIC
Warrington	-1.40	0.033	0.68	0.075	0.12	0.276	0.10	0.198	-1.27	0.248	-0.80	0.213	0.817	543.1
Warwick	-1.80	0.028	0.36	0.067	3.40	0.333	-1.36	0.209	-0.40	0.247	0.37	0.182	0.974	461.1
Wigan	-1.52	0.029	0.51	0.074	-0.79	0.275	0.88	0.194	-1.13	0.238	-0.50	0.213	0.861	583.7
Wokingham	-1.87	0.029	0.74	0.076	1.66	0.261	-3.03	0.189	-0.74	0.184	-1.28	0.147	0.981	529.8
Wolverhampton	-1.62	0.026	0.31	0.057	0.62	0.258	-0.11	0.168	-0.78	0.243	0.26	0.220	0.978	507.7
York	-1.41	0.053	0.55	0.058	2.16	0.259	0.12	0.164	-1.29	0.211	0.39	0.187	0.852	442.9

Appendix E: Competing destinations model goodness-of-fit and parameter estimates values

This appendix tabulates in full the AIC and R^2_{adj} goodness-of-fit statistics and the parameter estimates (with their standard errors) arising from global and origin-specific calibrations of the competing destinations migration destination choice model for all migrants aged 16+ years.

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Access.	SE	R2adj	AIC
Global	-1.30	0.002	0.88	0.005	0.91	0.019	-0.15	0.014	-0.42	0.016	-0.08	0.015	-0.41	0.005	0.830	91483.427
Aberdeen City	-1.51	0.111	1.25	0.061	3.22	0.314	0.82	0.255	-0.56	0.318	1.12	0.255	-0.68	0.061	0.962	404.548
Barking and Dagenham	-1.43	0.080	0.74	0.128	0.92	0.363	-2.12	0.325	-1.22	0.227	-0.81	0.263	-0.74	0.096	0.529	379.478
Barnsley	-2.15	0.058	0.35	0.068	-1.19	0.289	1.81	0.286	0.54	0.287	0.60	0.192	-0.91	0.075	0.931	515.691
Bath	-0.92	0.039	1.18	0.078	1.48	0.263	0.98	0.195	-1.29	0.225	0.29	0.199	-0.65	0.060	0.939	333.469
Birmingham	-1.39	0.025	1.11	0.042	0.84	0.128	0.19	0.089	-1.42	0.112	-0.37	0.091	-0.27	0.033	0.870	760.051
Blackburn	-1.34	0.053	0.78	0.093	-1.25	0.361	2.53	0.264	0.66	0.357	0.88	0.278	-0.67	0.087	0.662	445.078
Blackpool	-1.22	0.045	0.48	0.069	0.13	0.265	0.09	0.178	0.21	0.277	0.20	0.209	-0.06	0.059	0.728	447.961
Bolton	-1.62	0.032	0.66	0.063	-1.54	0.233	1.41	0.185	-0.81	0.219	-0.63	0.177	-0.72	0.069	0.936	575.044
Bournemouth	-0.73	0.036	0.98	0.077	0.33	0.252	0.82	0.198	-0.51	0.224	-0.04	0.200	-0.32	0.054	0.841	343.304
Bradford	-1.51	0.034	1.12	0.041	-0.17	0.186	1.68	0.130	-1.48	0.154	-0.20	0.137	-0.94	0.044	0.984	549.492
Brighton	-0.94	0.082	1.13	0.069	1.95	0.241	-0.09	0.234	-0.80	0.180	0.76	0.180	-0.45	0.052	0.774	387.854
Bristol	-0.85	0.031	0.99	0.047	1.02	0.147	0.88	0.106	-1.02	0.140	0.37	0.123	-0.45	0.034	0.713	537.718
Bury	-1.75	0.036	0.42	0.072	1.66	0.281	0.49	0.189	-1.93	0.230	0.73	0.234	-0.82	0.084	0.980	473.384
Cambridge	-0.45	0.065	0.92	0.059	2.08	0.210	0.62	0.157	-1.05	0.148	0.89	0.150	-0.49	0.045	0.655	482.155
Camden	-0.94	0.030	0.88	0.054	2.66	0.133	-0.96	0.123	-0.81	0.115	0.81	0.113	-0.43	0.045	0.972	583.992
Canterbury	-1.51	0.050	0.77	0.087	1.61	0.249	-0.72	0.216	0.08	0.207	1.11	0.198	-0.60	0.057	0.915	437.989
Cardiff	-1.16	0.029	0.74	0.048	2.55	0.199	-0.26	0.117	-0.23	0.192	1.24	0.164	-0.31	0.040	0.903	465.384
Carlisle	-1.57	0.130	0.64	0.083	0.74	0.366	0.64	0.305	-0.53	0.303	0.10	0.255	-0.33	0.072	0.545	318.614
Chelmsford	-1.58	0.067	0.96	0.094	1.41	0.257	-1.69	0.248	-1.01	0.200	-0.76	0.204	-0.54	0.061	0.618	492.215
Cheltenham	-1.70	0.041	0.67	0.076	1.63	0.326	-0.49	0.220	-1.45	0.298	0.17	0.236	-0.51	0.063	0.965	324.446
Chester	-1.02	0.058	0.89	0.068	1.86	0.259	0.41	0.176	-1.56	0.225	0.27	0.189	-0.40	0.065	0.784	359.045
Colchester	-1.18	0.051	0.77	0.080	1.67	0.239	-1.15	0.185	-1.06	0.188	-0.32	0.187	-0.31	0.054	0.680	443.554
Coventry	-1.48	0.025	0.40	0.044	2.86	0.201	-0.47	0.129	-0.18	0.186	1.18	0.153	-0.35	0.040	0.907	599.656
Croydon	-1.43	0.039	1.12	0.058	2.63	0.151	-1.72	0.157	-0.18	0.123	0.97	0.124	-0.77	0.046	0.871	652.794
Darlington	-1.34	0.075	0.85	0.091	0.64	0.370	1.11	0.312	-0.90	0.308	0.02	0.236	-0.49	0.085	0.798	255.466
Derby	-1.24	0.054	0.73	0.065	1.10	0.268	0.44	0.168	-0.80	0.244	0.76	0.214	-0.69	0.060	0.905	299.949

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Access.	SE	R2adj	AIC
Doncaster	-1.80	0.044	0.38	0.051	0.26	0.229	0.49	0.166	-0.51	0.234	0.25	0.180	-0.43	0.054	0.871	567.799
Dover	-1.52	0.049	0.86	0.118	1.08	0.314	-0.60	0.261	0.18	0.298	0.46	0.263	-0.51	0.080	0.948	360.439
Dudley	-1.65	0.042	0.24	0.055	0.48	0.297	0.08	0.208	0.02	0.319	1.16	0.237	-0.46	0.066	0.964	476.712
Dundee City	-1.34	0.094	1.14	0.070	1.47	0.403	0.88	0.280	-1.75	0.179	0.10	0.169	-0.58	0.073	0.984	287.826
Durham	-1.32	0.049	0.87	0.072	1.50	0.253	1.23	0.206	-1.14	0.238	0.40	0.200	-0.43	0.064	0.943	303.232
Ealing	-1.26	0.034	0.79	0.052	2.86	0.147	-1.55	0.139	0.73	0.122	1.48	0.119	-0.40	0.044	0.836	699.713
Edinburgh City	-0.93	0.055	1.06	0.049	2.80	0.206	0.25	0.155	-2.41	0.102	-0.01	0.100	-0.56	0.036	0.897	508.587
Exeter	-1.14	0.077	0.98	0.078	0.52	0.232	1.15	0.192	-1.00	0.231	0.19	0.208	-0.50	0.062	0.724	411.823
Glasgow City	-0.77	0.136	1.31	0.122	1.74	0.673	0.52	0.454	-1.66	0.344	-0.23	0.333	-0.70	0.111	0.876	172.402
Gloucester	-1.73	0.041	0.68	0.089	0.49	0.359	-0.30	0.257	-1.28	0.358	-0.19	0.282	-0.37	0.071	0.964	331.402
Greenwich	-0.91	0.043	0.91	0.067	0.95	0.181	0.22	0.185	0.05	0.149	0.86	0.150	-0.68	0.051	0.729	541.22
Guildford	-0.46	0.068	0.90	0.077	1.10	0.245	1.03	0.238	-0.49	0.190	0.49	0.182	-0.45	0.053	0.698	364.361
Hackney	-0.94	0.029	0.92	0.065	1.00	0.128	-0.19	0.117	-0.96	0.144	0.54	0.148	-0.54	0.051	0.985	511.634
Hammersmith and Fulham	-1.26	0.039	1.23	0.059	1.65	0.149	-0.31	0.165	0.65	0.126	1.19	0.139	-0.63	0.055	0.900	759.832
Harrogate	-1.91	0.044	0.63	0.043	0.22	0.234	1.67	0.170	-1.37	0.203	-0.23	0.176	-0.95	0.057	0.966	447.981
Harrow	-1.31	0.045	0.99	0.071	1.82	0.209	-0.93	0.200	0.45	0.164	0.84	0.163	-0.55	0.060	0.918	507.151
Ipswich	-1.43	0.076	0.92	0.123	1.80	0.389	-1.02	0.290	-1.57	0.281	-0.15	0.296	-0.51	0.084	0.871	200.555
Islington	-0.94	0.031	0.96	0.059	2.14	0.162	-0.48	0.161	-0.90	0.127	0.84	0.127	-0.63	0.046	0.981	551.547
Kensington and Chelsea	-1.29	0.033	0.58	0.068	2.82	0.170	-1.21	0.141	0.51	0.127	1.54	0.145	-0.53	0.062	0.966	689.227
Kings Lynn & W. Norfolk	-1.88	0.104	0.82	0.158	-0.48	0.446	0.92	0.337	-2.03	0.331	-0.59	0.374	-1.05	0.100	0.718	373.958
Kingston upon Hull	-1.50	0.073	1.03	0.054	0.97	0.257	0.87	0.178	-0.84	0.209	0.49	0.188	-0.56	0.059	0.844	349.744
Lambeth	-1.39	0.021	1.43	0.043	2.67	0.134	-1.58	0.124	-0.01	0.090	0.67	0.105	-0.61	0.038	0.958	732.721
Lancaster	-1.26	0.049	0.78	0.059	0.34	0.235	0.98	0.157	-0.62	0.215	0.10	0.174	-0.29	0.053	0.767	408.84
Leeds	-1.73	0.021	1.10	0.033	0.24	0.123	1.38	0.084	-2.00	0.091	-0.50	0.080	-0.70	0.031	0.904	832.655
Leicester	-0.88	0.047	0.95	0.046	1.66	0.204	0.34	0.124	-0.08	0.170	1.13	0.154	-0.35	0.042	0.669	495.005
Lincoln	-1.49	0.099	0.92	0.087	1.47	0.364	0.39	0.223	-0.10	0.322	0.76	0.282	-0.79	0.080	0.681	303.237
Liverpool	-1.12	0.035	0.99	0.045	1.11	0.176	0.86	0.124	-1.36	0.156	0.39	0.139	-0.54	0.045	0.799	516.315
Luton	-1.70	0.048	1.06	0.082	1.72	0.270	-1.79	0.193	-0.27	0.226	-0.11	0.218	-0.37	0.063	0.914	406.245
Macclesfield	-1.80	0.050	0.81	0.069	2.00	0.289	1.11	0.225	-2.52	0.227	-0.50	0.198	-0.95	0.083	0.972	455.501
Maidstone	-1.56	0.055	0.44	0.112	-0.01	0.233	-0.48	0.269	1.22	0.250	0.45	0.220	-0.37	0.064	0.919	481.139
Manchester	-1.31	0.015	0.56	0.032	1.99	0.137	0.57	0.104	-0.23	0.122	1.12	0.101	-0.17	0.038	0.943	828.266
Middlesbrough	-1.21	0.069	0.91	0.069	1.37	0.282	0.50	0.232	-1.02	0.258	0.30	0.221	-0.41	0.067	0.699	351.662
Milton Keynes	-1.46	0.049	0.76	0.072	1.09	0.256	-0.95	0.183	-0.50	0.207	-0.07	0.194	-0.33	0.055	0.796	432.913
Newbury	-2.41	0.049	0.62	0.095	0.82	0.261	-1.22	0.235	-1.07	0.252	-0.36	0.217	-0.44	0.066	0.869	675.101

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Access.	SE	R2adj	AIC
Newcastle upon Tyne	-0.95	0.031	1.04	0.042	1.11	0.176	0.72	0.136	-1.16	0.133	0.02	0.123	-0.27	0.036	0.876	456.663
Newport	-1.32	0.057	0.81	0.104	1.98	0.370	-0.26	0.260	-0.71	0.381	0.62	0.306	-0.41	0.084	0.934	360.355
Northampton	-1.52	0.046	0.76	0.068	1.44	0.261	-0.82	0.170	-0.42	0.226	0.30	0.206	-0.46	0.054	0.879	385.052
Norwich	-2.63	0.119	1.00	0.143	1.15	0.463	-1.05	0.328	-1.27	0.395	-0.78	0.376	-0.49	0.109	0.706	415.269
Nottingham	-1.24	0.040	0.87	0.047	1.83	0.199	0.48	0.120	-0.83	0.173	0.86	0.156	-0.56	0.044	0.902	403.556
Oldham	-1.98	0.047	0.39	0.082	0.97	0.332	1.65	0.262	-1.23	0.261	1.54	0.268	-1.34	0.102	0.966	512.709
Oxford	-0.76	0.052	0.93	0.052	2.96	0.192	-0.05	0.144	-1.18	0.137	1.11	0.133	-0.47	0.041	0.746	504.193
Peterborough	-1.41	0.077	0.80	0.083	0.41	0.296	-0.02	0.200	-0.93	0.245	-0.09	0.237	-0.58	0.064	0.690	294.631
Plymouth	-1.11	0.058	0.98	0.060	-0.52	0.185	1.14	0.148	-0.52	0.178	-0.24	0.152	-0.45	0.046	0.716	463.69
Poole	-1.21	0.058	0.84	0.147	1.13	0.467	-1.25	0.338	-1.35	0.433	-1.02	0.366	-0.14	0.102	0.968	165.09
Portsmouth	-0.81	0.050	1.08	0.068	-0.25	0.212	0.77	0.192	-0.63	0.187	0.01	0.167	-0.66	0.047	0.689	434.673
Preston	-1.40	0.050	0.65	0.076	0.11	0.293	1.18	0.191	-1.06	0.278	0.30	0.223	-0.63	0.068	0.823	403.018
Reading	-1.51	0.028	1.29	0.065	0.27	0.182	-1.00	0.154	-2.43	0.134	-2.00	0.124	-0.61	0.046	0.926	743.5
Rochdale	-1.93	0.048	0.51	0.075	0.19	0.297	0.90	0.231	-2.06	0.244	0.02	0.231	-1.08	0.099	0.941	544.174
Rochester upon Medway	-1.87	0.047	0.88	0.107	1.04	0.244	-1.73	0.244	0.06	0.247	0.18	0.220	-0.61	0.068	0.903	507.314
Rotherham	-2.01	0.039	0.40	0.068	-0.53	0.284	0.38	0.216	0.94	0.272	0.86	0.214	-0.70	0.066	0.979	539.24
Salford	-1.70	0.031	0.86	0.060	-1.87	0.233	1.53	0.191	-1.00	0.182	-0.67	0.167	-0.73	0.068	0.900	715.488
Scarborough	-2.15	0.073	1.06	0.061	-0.63	0.270	1.05	0.213	-0.51	0.301	-0.99	0.201	-0.31	0.068	0.877	368.201
Sheffield	-1.85	0.025	0.58	0.033	1.59	0.156	0.86	0.094	-1.13	0.130	1.24	0.117	-0.87	0.039	0.921	742.369
Southampton	-1.06	0.054	0.87	0.072	1.43	0.226	-0.07	0.201	-0.62	0.204	0.64	0.169	-0.58	0.053	0.850	358.828
Southwark	-1.16	0.023	1.20	0.056	1.27	0.131	-0.26	0.132	0.23	0.122	0.82	0.122	-0.66	0.045	0.963	627.375
St Albans	-1.74	0.046	0.54	0.090	1.33	0.239	-1.72	0.202	0.56	0.201	0.95	0.199	-0.80	0.065	0.841	537.198
Stafford	-1.74	0.073	0.79	0.072	-1.84	0.267	0.97	0.211	-2.01	0.276	-1.65	0.195	-0.68	0.070	0.944	437.754
Stirling	-1.26	0.102	1.17	0.105	2.42	0.561	0.82	0.389	-0.97	0.289	0.26	0.262	-0.74	0.087	0.987	169.919
Stockport	-1.68	0.029	0.88	0.056	2.00	0.247	1.07	0.183	-1.66	0.173	0.39	0.172	-0.92	0.062	0.990	524.691
Stoke-on-Trent	-2.11	0.069	0.35	0.078	0.11	0.347	-0.18	0.208	-2.65	0.271	-1.07	0.242	-0.83	0.082	0.801	561.619
Stratford-on-Avon	-2.03	0.044	0.58	0.071	2.59	0.365	-0.97	0.244	-1.61	0.321	-0.11	0.263	-0.38	0.066	0.957	453.283
Sunderland	-1.23	0.047	0.91	0.060	0.27	0.260	0.84	0.203	-0.80	0.201	-0.10	0.172	-0.24	0.053	0.938	382.6
Swansea	-1.46	0.074	0.95	0.076	1.98	0.277	0.41	0.187	-0.20	0.290	0.83	0.239	-0.38	0.058	0.926	299.138
Thamesdown	-1.42	0.060	0.84	0.074	0.50	0.269	-0.11	0.202	-1.15	0.232	-0.45	0.182	-0.65	0.056	0.774	395.062
Tower Hamlets	-0.77	0.040	0.57	0.076	-0.89	0.159	0.78	0.180	-0.41	0.164	0.24	0.157	-0.18	0.053	0.903	561.962
Trafford	-1.59	0.026	0.68	0.060	2.47	0.220	0.26	0.159	-2.07	0.176	0.39	0.182	-0.76	0.065	0.985	565.676
Wakefield	-2.10	0.052	0.54	0.053	0.53	0.263	0.91	0.190	-0.39	0.234	0.50	0.199	-1.10	0.064	0.953	606.173
Walsall	-1.90	0.043	0.51	0.050	1.16	0.342	-0.25	0.222	-0.77	0.340	0.98	0.248	-0.51	0.068	0.993	406.494

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Access.	SE	R2adj	AIC
Warrington	-1.65	0.045	0.83	0.076	-0.05	0.279	0.72	0.214	-1.86	0.250	-0.88	0.202	-0.69	0.075	0.854	521.845
Warwick	-1.89	0.032	0.38	0.065	3.54	0.333	-1.31	0.206	-0.73	0.256	0.50	0.185	-0.36	0.062	0.974	464.357
Wigan	-1.76	0.040	0.71	0.075	-0.87	0.281	1.44	0.207	-2.12	0.249	-0.77	0.206	-0.74	0.073	0.885	565.78
Wokingham	-1.81	0.028	1.02	0.076	1.82	0.265	-2.30	0.200	-1.11	0.184	-0.89	0.146	-0.64	0.054	0.984	514.847
Wolverhampton	-1.56	0.041	0.35	0.062	0.57	0.260	-0.12	0.169	-0.62	0.255	0.23	0.221	0.12	0.058	0.983	483.909
York	-1.69	0.054	0.67	0.054	2.65	0.253	1.10	0.172	-1.32	0.212	1.28	0.193	-0.81	0.060	0.911	393.476

Appendix F: Discrete nested logit model goodness-of-fit and parameter estimates values

This appendix tabulates in full the AIC and R^2_{adj} goodness-of-fit statistics and the parameter estimates (with their standard errors) arising from global and origin-specific calibrations of the discrete nested logit migration destination choice model for all migrants aged 16+ years.

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Reg. util.	SE	R2adj	AIC
Global	-1.24	0.002	0.81	0.006	0.98	0.020	-0.57	0.014	-0.29	0.016	-0.24	0.015	-0.27	0.012	0.814	92408.1
Aberdeen City	-2.35	0.107	1.20	0.063	3.29	0.342	0.54	0.259	0.24	0.313	1.18	0.268	0.94	0.243	0.939	452.8
Barking and Dagenham	-0.94	0.088	0.53	0.130	0.18	0.365	-2.27	0.335	-1.00	0.247	-1.48	0.263	1.52	0.294	0.499	385.4
Barnsley	-1.72	0.059	0.42	0.069	-0.93	0.294	1.12	0.275	0.78	0.269	0.54	0.197	-0.67	0.200	0.930	517.4
Bath	-1.18	0.046	0.91	0.080	1.20	0.293	0.34	0.205	-0.84	0.237	-0.06	0.212	0.40	0.245	0.914	367.3
Birmingham	-1.22	0.019	0.99	0.039	0.64	0.132	0.14	0.101	-1.11	0.106	-0.45	0.092	-0.14	0.086	0.879	752.5
Blackburn	-1.17	0.043	0.41	0.090	-1.24	0.341	1.75	0.250	1.74	0.334	1.19	0.278	0.54	0.118	0.723	425.6
Blackpool	-1.22	0.046	0.46	0.065	0.23	0.273	-0.07	0.187	0.16	0.283	0.16	0.213	0.23	0.173	0.722	450.0
Bolton	-1.52	0.026	0.63	0.063	-1.47	0.223	0.70	0.166	-0.83	0.223	-1.06	0.190	-0.74	0.079	0.916	602.8
Bournemouth	-0.81	0.034	0.86	0.078	0.28	0.263	0.41	0.198	-0.35	0.227	-0.24	0.201	0.00	0.190	0.792	369.6
Bradford	-1.13	0.033	1.38	0.039	-0.46	0.188	0.58	0.138	-1.66	0.151	-1.44	0.126	0.12	0.107	0.946	669.4
Brighton	-0.83	0.106	0.99	0.074	1.78	0.242	-0.53	0.231	-0.50	0.182	0.45	0.180	-0.31	0.210	0.742	400.9
Bristol	-0.92	0.031	0.82	0.047	0.96	0.153	0.46	0.103	-0.66	0.144	0.22	0.126	-0.19	0.116	0.659	554.8
Bury	-1.55	0.024	0.27	0.070	1.89	0.278	-0.60	0.169	-1.62	0.238	0.52	0.244	-0.44	0.078	0.973	500.1
Cambridge	-0.39	0.080	0.71	0.058	1.68	0.211	0.45	0.162	-0.82	0.153	0.51	0.149	0.42	0.131	0.590	499.3
Camden	-0.91	0.030	0.68	0.051	2.16	0.147	-0.83	0.132	-0.77	0.116	0.37	0.115	-0.64	0.105	0.973	580.9
Canterbury	-1.77	0.054	0.41	0.083	1.56	0.247	-1.35	0.207	0.45	0.215	0.77	0.210	-0.55	0.149	0.893	460.9
Cardiff	-1.17	0.037	0.69	0.050	2.55	0.207	-0.53	0.123	0.05	0.193	1.16	0.168	-0.28	0.123	0.894	474.5
Carlisle	-1.58	0.157	0.56	0.083	1.09	0.366	-0.12	0.270	-0.65	0.313	-0.19	0.263	-0.30	0.184	0.569	313.1
Chelmsford	-1.72	0.084	0.61	0.095	1.48	0.247	-2.35	0.239	-0.68	0.211	-1.13	0.209	-0.31	0.173	0.532	512.4
Cheltenham	-1.65	0.053	0.55	0.079	1.90	0.354	-0.87	0.229	-0.70	0.295	0.33	0.250	0.26	0.221	0.948	364.3
Chester	-0.79	0.052	0.83	0.068	1.86	0.263	-0.03	0.177	-1.12	0.218	0.24	0.196	-0.09	0.158	0.737	378.4
Colchester	-1.50	0.069	0.59	0.073	1.40	0.237	-1.13	0.183	-1.09	0.195	-0.67	0.185	-1.15	0.177	0.718	430.8
Coventry	-1.43	0.027	0.30	0.051	2.62	0.210	-0.52	0.131	0.28	0.180	1.14	0.156	0.23	0.096	0.910	596.6
Croydon	-1.25	0.044	0.81	0.059	2.47	0.157	-2.06	0.156	0.30	0.127	0.68	0.131	-0.06	0.120	0.801	695.5
Darlington	-1.29	0.083	0.67	0.090	0.80	0.382	0.17	0.280	-0.83	0.313	-0.27	0.249	0.35	0.245	0.759	273.0
Derby	-1.14	0.070	0.54	0.069	0.78	0.265	-0.10	0.170	-0.20	0.241	0.57	0.216	0.53	0.149	0.767	388.5

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Reg. util.	SE	R2adj	AIC
Doncaster	-1.69	0.038	0.44	0.054	0.13	0.222	-0.21	0.164	-0.38	0.220	-0.11	0.176	-0.99	0.108	0.898	544.8
Dover	-1.92	0.068	0.62	0.112	0.71	0.319	-0.70	0.256	0.10	0.317	-0.18	0.283	-1.26	0.223	0.952	351.9
Dudley	-1.49	0.052	0.38	0.075	0.54	0.306	0.03	0.223	0.79	0.301	1.23	0.241	0.14	0.175	0.951	508.1
Dundee City	-1.72	0.087	1.18	0.073	1.14	0.420	0.31	0.324	-1.84	0.180	-0.76	0.131	0.70	0.311	0.978	321.2
Durham	-1.34	0.051	0.78	0.073	1.42	0.256	0.50	0.203	-1.13	0.244	-0.02	0.202	0.42	0.157	0.929	325.4
Ealing	-1.10	0.028	0.62	0.050	2.74	0.150	-1.50	0.141	1.02	0.121	1.38	0.122	0.22	0.127	0.840	697.4
Edinburgh City	-1.04	0.065	0.90	0.049	2.57	0.205	-0.59	0.149	-2.82	0.100	-0.92	0.084	-0.11	0.147	0.840	552.1
Exeter	-1.52	0.067	0.84	0.079	0.30	0.231	0.67	0.185	-0.89	0.245	-0.22	0.209	-0.28	0.206	0.691	422.8
Glasgow City	-0.95	0.144	1.20	0.122	1.53	0.681	-0.10	0.474	-2.07	0.348	-1.16	0.307	0.39	0.308	0.776	231.0
Gloucester	-1.76	0.041	0.69	0.090	-0.24	0.397	0.04	0.278	-0.98	0.354	-0.70	0.311	-0.93	0.190	0.972	309.0
Greenwich	-0.70	0.041	0.61	0.066	0.59	0.181	0.11	0.190	0.39	0.155	0.54	0.155	0.00	0.138	0.649	566.9
Guildford	-0.42	0.077	0.68	0.078	0.90	0.249	0.70	0.239	-0.13	0.195	0.27	0.187	0.06	0.149	0.620	387.1
Hackney	-0.86	0.028	0.68	0.061	0.76	0.130	-0.35	0.118	-1.02	0.145	0.02	0.143	-0.59	0.132	0.982	526.6
Hammersmith and Fulham	-1.06	0.032	0.95	0.054	1.44	0.154	-0.08	0.164	0.93	0.128	1.08	0.148	0.55	0.117	0.894	766.1
Harrogate	-1.89	0.050	0.67	0.047	-0.50	0.234	0.36	0.158	-2.15	0.195	-1.92	0.174	-1.08	0.121	0.928	522.6
Harrow	-1.14	0.043	0.77	0.069	1.61	0.206	-0.98	0.201	0.96	0.161	0.69	0.166	-0.08	0.123	0.892	534.5
Ipswich	-1.57	0.106	0.68	0.124	1.71	0.383	-1.70	0.276	-1.40	0.303	-0.59	0.301	-0.07	0.248	0.817	235.1
Islington	-0.98	0.036	0.71	0.056	2.14	0.161	-1.02	0.160	-0.70	0.130	0.40	0.129	-0.70	0.092	0.964	615.9
Kensington and Chelsea	-1.06	0.028	0.32	0.060	2.94	0.174	-1.03	0.152	0.90	0.119	1.75	0.151	0.26	0.078	0.968	682.6
Kings Lynn & W. Norfolk	-2.03	0.130	0.45	0.165	-0.65	0.404	-0.71	0.328	-2.12	0.357	-1.86	0.354	1.16	0.232	0.548	420.8
Kingston upon Hull	-1.20	0.076	0.90	0.056	0.92	0.261	0.14	0.189	-0.65	0.210	0.08	0.186	0.00	0.179	0.771	387.6
Lambeth	-1.32	0.021	1.26	0.043	2.26	0.133	-1.70	0.128	-0.01	0.092	0.05	0.102	-0.55	0.096	0.952	745.2
Lancaster	-1.15	0.050	0.67	0.057	0.32	0.235	0.66	0.154	-0.34	0.214	0.04	0.182	0.02	0.120	0.745	418.1
Leeds	-1.49	0.018	1.08	0.035	0.08	0.123	0.43	0.092	-1.84	0.093	-1.25	0.082	0.48	0.063	0.928	804.5
Leicester	-0.74	0.050	0.90	0.051	1.45	0.203	0.13	0.126	0.23	0.167	0.96	0.154	0.08	0.112	0.666	496.0
Lincoln	-1.10	0.115	0.74	0.094	0.67	0.350	-0.14	0.231	0.03	0.319	0.13	0.276	0.01	0.232	0.521	343.4
Liverpool	-0.89	0.030	0.87	0.046	1.22	0.177	0.06	0.117	-0.88	0.155	0.16	0.143	0.22	0.077	0.707	553.9
Luton	-1.80	0.056	0.79	0.082	1.28	0.276	-1.74	0.197	-0.19	0.228	-0.38	0.215	-0.89	0.168	0.934	380.7
Macclesfield	-1.43	0.034	0.80	0.074	2.54	0.293	-0.35	0.184	-1.97	0.231	-0.96	0.214	-0.35	0.111	0.962	485.1
Maidstone	-1.65	0.062	0.07	0.098	0.10	0.230	-0.77	0.257	1.50	0.247	0.29	0.233	-0.46	0.145	0.928	470.6
Manchester	-1.04	0.014	0.71	0.032	2.11	0.135	0.56	0.096	-0.22	0.116	1.12	0.101	-1.15	0.059	0.970	766.8
Middlesbrough	-1.02	0.075	0.79	0.067	1.57	0.291	-0.12	0.206	-0.96	0.268	0.12	0.228	-0.60	0.244	0.734	339.6
Milton Keynes	-1.49	0.053	0.58	0.073	0.70	0.262	-0.95	0.188	-0.36	0.206	-0.36	0.196	-0.48	0.138	0.782	439.3
Newbury	-2.64	0.052	0.32	0.082	0.36	0.264	-1.34	0.226	-0.55	0.228	-0.40	0.215	-1.17	0.145	0.897	651.1

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Reg. util.	SE	R2adj	AIC
Newcastle upon Tyne	-0.93	0.034	0.96	0.042	1.15	0.183	0.33	0.130	-1.21	0.137	-0.21	0.121	-0.08	0.081	0.876	456.0
Newport	-1.48	0.061	0.67	0.103	1.87	0.392	-0.64	0.257	-0.47	0.390	0.42	0.324	-0.07	0.205	0.949	335.7
Northampton	-1.58	0.062	0.60	0.070	1.23	0.264	-0.99	0.174	-0.20	0.228	0.08	0.208	-0.39	0.154	0.828	420.1
Norwich	-2.79	0.132	0.81	0.146	1.00	0.458	-1.53	0.317	-1.04	0.420	-1.14	0.394	0.21	0.276	0.736	404.5
Nottingham	-1.10	0.046	0.76	0.049	1.43	0.196	0.05	0.126	-0.33	0.171	0.53	0.160	0.31	0.106	0.814	466.8
Oldham	-1.53	0.031	0.01	0.077	1.48	0.323	-0.10	0.204	-0.44	0.264	1.54	0.272	0.10	0.111	0.927	588.4
Oxford	-0.70	0.062	0.76	0.054	2.74	0.198	-0.37	0.145	-0.83	0.139	0.82	0.136	0.14	0.117	0.697	521.5
Peterborough	-1.16	0.102	0.63	0.088	0.20	0.288	-0.58	0.206	-0.61	0.256	-0.49	0.234	0.62	0.210	0.599	320.2
Plymouth	-1.28	0.057	0.83	0.060	-0.16	0.195	0.51	0.141	-0.31	0.186	-0.26	0.160	-1.06	0.171	0.698	470.0
Poole	-1.29	0.058	0.76	0.147	0.81	0.499	-1.21	0.343	-1.36	0.431	-1.27	0.363	-0.47	0.298	0.972	150.8
Portsmouth	-1.02	0.054	0.80	0.069	-0.17	0.223	-0.28	0.191	-0.11	0.195	-0.28	0.174	0.05	0.107	0.504	480.9
Preston	-1.20	0.043	0.45	0.077	0.17	0.297	0.42	0.201	-0.28	0.275	0.34	0.233	0.40	0.128	0.737	442.2
Reading	-1.45	0.029	1.10	0.070	0.03	0.195	-1.61	0.172	-2.31	0.141	-2.66	0.121	0.04	0.163	0.916	755.8
Rochdale	-1.52	0.027	0.24	0.075	0.41	0.285	-0.26	0.195	-1.53	0.252	-0.17	0.243	0.12	0.104	0.916	579.5
Rochester upon Medway	-1.98	0.049	0.33	0.097	1.16	0.229	-2.10	0.230	0.84	0.244	0.12	0.226	-0.75	0.169	0.902	508.7
Rotherham	-2.02	0.042	0.31	0.070	-0.23	0.304	0.16	0.234	1.09	0.257	1.11	0.219	1.94	0.226	0.984	513.1
Salford	-1.51	0.022	0.75	0.058	-1.62	0.225	0.78	0.169	-0.89	0.179	-0.83	0.177	-0.42	0.092	0.885	729.1
Scarborough	-2.04	0.083	0.93	0.061	-0.52	0.289	0.69	0.220	-0.39	0.317	-0.92	0.222	0.42	0.139	0.898	350.2
Sheffield	-1.54	0.020	0.45	0.036	0.84	0.147	0.09	0.100	-0.36	0.126	0.73	0.115	0.71	0.081	0.896	769.4
Southampton	-1.33	0.054	0.60	0.071	1.35	0.246	-0.79	0.198	-0.24	0.209	0.33	0.182	-0.36	0.134	0.801	386.7
Southwark	-1.04	0.021	0.87	0.055	1.17	0.132	-0.45	0.138	0.50	0.124	0.67	0.125	0.61	0.104	0.958	639.8
St Albans	-1.70	0.046	0.02	0.080	0.65	0.227	-1.65	0.205	0.83	0.203	0.26	0.204	-1.33	0.153	0.807	556.1
Stafford	-1.37	0.066	0.77	0.092	-2.68	0.232	0.66	0.202	-1.45	0.272	-2.18	0.195	-0.41	0.175	0.884	509.0
Stirling	-1.50	0.093	1.30	0.102	2.61	0.551	-0.56	0.360	-1.27	0.300	-0.80	0.239	1.31	0.259	0.977	222.1
Stockport	-1.42	0.021	0.84	0.058	2.86	0.248	-0.60	0.153	-1.84	0.173	-0.23	0.169	-0.30	0.075	0.981	586.3
Stoke-on-Trent	-1.81	0.062	0.26	0.081	0.59	0.334	-1.25	0.191	-2.22	0.263	-1.42	0.240	-0.72	0.143	0.782	570.9
Stratford-on-Avon	-1.98	0.040	0.29	0.087	2.09	0.372	-0.85	0.243	-1.04	0.310	-0.14	0.266	-0.92	0.177	0.964	435.2
Sunderland	-1.25	0.049	0.90	0.061	0.22	0.264	0.19	0.193	-0.66	0.206	-0.42	0.180	0.96	0.155	0.938	382.3
Swansea	-1.70	0.083	0.84	0.078	1.70	0.288	0.19	0.213	0.05	0.294	0.60	0.248	0.18	0.305	0.911	317.8
Thamesdown	-1.47	0.071	0.67	0.078	0.58	0.277	-0.77	0.204	-0.51	0.237	-0.54	0.194	0.31	0.163	0.669	433.0
Tower Hamlets	-0.71	0.037	0.46	0.071	-0.90	0.156	0.50	0.182	-0.31	0.164	0.18	0.156	-0.79	0.144	0.901	563.5
Trafford	-1.39	0.021	0.51	0.059	2.85	0.220	-0.79	0.133	-2.03	0.184	0.11	0.190	-0.01	0.077	0.982	584.0
Wakefield	-1.74	0.044	0.86	0.048	-0.12	0.253	-0.22	0.199	-0.95	0.208	-0.90	0.171	-0.78	0.149	0.930	645.4
Walsall	-1.64	0.042	0.78	0.057	0.98	0.367	0.00	0.240	-0.11	0.340	1.02	0.270	-0.38	0.218	0.985	484.8

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Reg. util.	SE	R2adj	AIC
Warrington	-1.33	0.040	0.64	0.076	0.10	0.275	0.24	0.204	-1.16	0.251	-0.67	0.214	0.38	0.112	0.822	541.3
Warwick	-1.79	0.052	0.37	0.075	3.43	0.345	-1.37	0.210	-0.40	0.248	0.37	0.183	0.07	0.205	0.974	463.2
Wigan	-1.55	0.029	0.51	0.072	-0.69	0.269	0.61	0.191	-1.45	0.240	-0.70	0.213	-0.39	0.057	0.900	552.7
Wokingham	-1.96	0.031	0.66	0.072	1.27	0.271	-2.48	0.202	-0.87	0.187	-1.52	0.150	-1.23	0.147	0.983	518.2
Wolverhampton	-1.61	0.027	0.34	0.067	0.67	0.264	-0.10	0.170	-0.77	0.244	0.30	0.225	0.12	0.117	0.980	503.1
York	-1.26	0.067	0.54	0.058	2.21	0.260	0.21	0.166	-1.38	0.213	0.41	0.185	-0.65	0.180	0.855	441.2

Appendix G: Weighted nested logit model goodness-of-fit and parameter estimates values

This appendix tabulates in full the AIC and R^2_{adj} goodness-of-fit statistics and the parameter estimates (with their standard errors) arising from global and origin-specific calibrations of the weighted nested logit migration destination choice model for all migrants aged 16+ years.

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Reg. util.	SE	R2adj	AIC
Global	-1.24	0.002	0.79	0.006	1.01	0.020	-0.67	0.018	-0.29	0.016	-0.29	0.015	-0.21	0.014	0.812	92,519.2
Aberdeen City	-2.17	0.095	1.20	0.062	3.13	0.321	-0.33	0.290	0.10	0.308	0.65	0.262	1.78	0.280	0.946	440.3
Barking and Dagenham	-0.70	0.114	0.59	0.132	0.97	0.390	-2.55	0.347	-1.20	0.264	-1.22	0.284	3.08	0.516	0.584	367.0
Barnsley	-2.26	0.070	0.42	0.073	-0.37	0.285	-0.16	0.285	1.31	0.269	0.55	0.198	3.29	0.377	0.942	497.8
Bath	-0.97	0.059	1.02	0.078	1.78	0.300	0.02	0.195	-1.01	0.244	0.11	0.211	-1.22	0.356	0.901	381.5
Birmingham	-1.16	0.019	1.04	0.040	0.50	0.132	0.60	0.125	-1.16	0.107	-0.38	0.092	-0.88	0.143	0.881	751.2
Blackburn	-1.44	0.081	0.56	0.091	-1.06	0.339	1.13	0.293	1.56	0.330	0.83	0.284	1.77	0.362	0.652	448.1
Blackpool	-1.32	0.051	0.46	0.064	0.06	0.271	0.37	0.190	0.23	0.272	0.34	0.214	-2.39	0.546	0.769	431.6
Bolton	-1.88	0.039	0.64	0.062	-1.34	0.218	0.42	0.153	0.09	0.212	-0.50	0.184	-2.54	0.159	0.965	514.4
Bournemouth	-0.74	0.040	0.88	0.077	0.50	0.259	0.09	0.206	-0.36	0.226	-0.16	0.201	-0.75	0.219	0.837	345.6
Bradford	-1.31	0.052	1.35	0.041	-0.40	0.183	0.19	0.158	-1.49	0.155	-1.47	0.125	1.03	0.220	0.952	658.4
Brighton	-0.42	0.124	1.08	0.072	1.67	0.239	-0.39	0.232	-0.39	0.180	0.50	0.178	-1.50	0.284	0.757	395.0
Bristol	-0.70	0.047	0.78	0.046	1.40	0.170	0.25	0.108	-0.66	0.143	0.48	0.133	-1.29	0.203	0.677	549.5
Bury	-1.85	0.044	0.43	0.074	1.88	0.272	-1.11	0.177	-1.49	0.232	0.18	0.246	-2.14	0.230	0.977	485.1
Cambridge	0.03	0.109	0.79	0.060	1.85	0.213	0.22	0.168	-0.77	0.151	0.53	0.148	-0.75	0.235	0.586	500.4
Camden	-0.96	0.041	0.74	0.052	2.17	0.162	-0.83	0.144	-0.66	0.120	0.45	0.114	-0.92	0.219	0.972	583.2
Canterbury	-1.97	0.070	0.37	0.083	1.61	0.254	-1.11	0.215	0.58	0.211	0.92	0.209	-1.26	0.222	0.928	420.9
Cardiff	-1.11	0.047	0.66	0.049	2.76	0.228	-0.57	0.123	-0.03	0.196	1.24	0.172	-0.79	0.273	0.899	469.3
Carlisle	-2.26	0.244	0.66	0.087	0.64	0.373	-0.15	0.269	-0.73	0.317	-0.56	0.283	-1.42	0.359	0.649	292.8
Chelmsford	-2.06	0.140	0.56	0.096	1.49	0.245	-2.31	0.234	-0.66	0.216	-1.10	0.212	-1.14	0.324	0.571	503.7
Cheltenham	-1.87	0.059	0.66	0.083	1.43	0.354	-0.45	0.239	-0.62	0.293	0.09	0.251	-1.06	0.249	0.961	334.6
Chester	-1.04	0.087	0.86	0.070	2.01	0.260	-0.54	0.211	-1.18	0.216	0.02	0.201	0.91	0.278	0.749	373.9
Colchester	-1.55	0.107	0.57	0.078	1.42	0.242	-1.10	0.207	-0.92	0.201	-0.49	0.193	-1.02	0.279	0.722	429.4
Coventry	-1.38	0.056	0.36	0.045	2.74	0.204	-0.53	0.180	0.29	0.181	1.16	0.160	-0.11	0.332	0.908	599.1
Croydon	-1.40	0.060	0.79	0.058	2.27	0.162	-1.88	0.163	0.32	0.127	0.58	0.132	-0.89	0.260	0.809	691.8
Darlington	-1.46	0.090	0.81	0.097	0.37	0.383	0.01	0.279	-0.64	0.313	-0.65	0.259	1.93	0.404	0.768	269.4
Derby	-0.93	0.089	0.61	0.069	0.61	0.265	0.20	0.229	-0.26	0.245	0.56	0.220	-0.28	0.336	0.718	407.3

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Reg. util.	SE	R2adj	AIC
Doncaster	-2.47	0.068	0.47	0.055	0.02	0.222	-0.38	0.152	0.23	0.219	0.02	0.176	-3.37	0.223	0.951	471.9
Dover	-2.05	0.080	0.53	0.115	0.40	0.344	-0.16	0.306	0.37	0.308	-0.01	0.282	-1.89	0.304	0.968	312.5
Dudley	-1.54	0.046	0.50	0.058	0.52	0.301	-0.22	0.240	0.87	0.299	1.04	0.248	1.06	0.405	0.957	494.5
Dundee City	-1.46	0.090	1.17	0.072	1.13	0.391	-0.53	0.333	-1.39	0.189	-0.80	0.129	2.36	0.380	0.984	289.8
Durham	-1.41	0.053	0.80	0.075	1.24	0.253	0.29	0.208	-0.95	0.248	-0.19	0.208	1.17	0.229	0.934	318.0
Ealing	-1.10	0.027	0.66	0.049	2.36	0.159	-1.16	0.150	1.14	0.123	1.31	0.123	-1.08	0.195	0.847	693.2
Edinburgh City	-1.36	0.071	1.02	0.052	2.61	0.205	-1.02	0.159	-2.68	0.100	-1.06	0.087	-1.37	0.185	0.859	539.3
Exeter	-1.29	0.085	0.86	0.076	0.46	0.241	0.79	0.186	-1.05	0.247	-0.05	0.209	-2.22	0.473	0.685	424.9
Glasgow City	-1.21	0.191	1.26	0.122	1.59	0.675	-0.65	0.490	-2.22	0.359	-1.37	0.329	-0.69	0.523	0.796	221.8
Gloucester	-2.09	0.074	0.78	0.096	-0.23	0.385	0.07	0.272	-0.51	0.349	-0.43	0.294	-1.89	0.317	0.979	279.4
Greenwich	-0.62	0.054	0.63	0.066	0.61	0.184	0.06	0.192	0.44	0.156	0.56	0.154	-0.80	0.335	0.653	565.8
Guildford	-0.17	0.108	0.76	0.081	0.98	0.246	0.66	0.239	-0.10	0.192	0.30	0.186	-0.76	0.268	0.621	386.8
Hackney	-0.99	0.041	0.73	0.062	0.56	0.140	-0.05	0.142	-0.74	0.156	0.10	0.144	-1.37	0.249	0.984	514.8
Hammersmith and Fulham	-1.26	0.041	0.98	0.055	1.03	0.157	0.11	0.168	0.74	0.129	0.83	0.148	2.64	0.277	0.896	764.1
Harrogate	-2.32	0.076	0.65	0.045	-0.64	0.225	0.29	0.153	-1.15	0.202	-1.48	0.164	.-2.57	0.233	0.950	487.7
Harrow	-1.08	0.041	0.91	0.069	0.94	0.216	-0.30	0.214	1.16	0.164	0.54	0.167	-4.53	0.466	0.923	500.3
Ipswich	-1.56	0.142	0.68	0.125	1.71	0.392	-1.71	0.309	-1.39	0.301	-0.59	0.301	-0.04	0.387	0.816	235.5
Islington	-1.04	0.037	0.75	0.057	1.35	0.176	-0.20	0.178	-0.47	0.134	0.33	0.129	-1.80	0.190	0.982	544.8
Kensington and Chelsea	-1.16	0.027	0.36	0.061	2.50	0.185	-0.95	0.148	1.02	0.122	1.50	0.153	-1.15	0.219	0.970	675.6
Kings Lynn & W. Norfolk	-2.23	0.204	0.31	0.165	-1.06	0.422	-0.29	0.389	-2.07	0.359	-1.81	0.362	0.19	0.484	0.428	444.1
Kingston upon Hull	-1.40	0.121	0.92	0.057	1.01	0.255	-0.16	0.212	-0.60	0.211	0.04	0.187	0.71	0.334	0.780	383.9
Lambeth	-1.45	0.025	1.22	0.043	1.70	0.141	-1.10	0.139	0.18	0.093	0.02	0.100	-1.97	0.170	0.957	735.5
Lancaster	-1.43	0.085	0.76	0.062	0.36	0.232	0.23	0.181	-0.44	0.211	-0.20	0.188	1.15	0.291	0.769	408.1
Leeds	-1.70	0.026	1.05	0.036	0.29	0.125	-0.04	0.101	-1.43	0.091	-1.16	0.079	1.77	0.135	0.940	785.9
Leicester	-0.71	0.104	0.91	0.048	1.44	0.203	0.17	0.150	0.23	0.167	0.96	0.159	-0.05	0.315	0.671	494.5
Lincoln	-0.97	0.173	0.72	0.091	0.62	0.349	0.09	0.323	0.04	0.319	0.20	0.288	-0.52	0.565	0.515	344.6
Liverpool	-1.19	0.056	0.91	0.046	1.33	0.175	-0.46	0.142	-0.86	0.152	-0.01	0.144	1.42	0.208	0.755	536.0
Luton	-1.92	0.083	0.84	0.084	1.30	0.282	-1.69	0.217	-0.13	0.231	-0.36	0.218	-1.00	0.253	0.933	381.7
Macclesfield	-1.88	0.084	0.73	0.072	2.11	0.278	-0.71	0.181	-1.86	0.226	-1.00	0.202	-2.04	0.315	0.965	476.8
Maidstone	-2.18	0.085	-0.01	0.097	0.27	0.244	-0.49	0.256	0.91	0.249	0.06	0.225	-2.45	0.247	0.967	392.3
Manchester	-1.59	0.022	0.61	0.032	1.99	0.130	-0.56	0.107	-0.16	0.116	0.58	0.105	2.31	0.137	0.966	776.7
Middlesbrough	-1.10	0.068	0.76	0.067	1.67	0.302	-0.15	0.208	-1.23	0.271	0.08	0.225	-0.93	0.366	0.745	335.3
Milton Keynes	-1.81	0.091	0.58	0.073	0.62	0.267	-0.51	0.221	-0.24	0.208	-0.20	0.196	-1.38	0.268	0.820	420.1
Newbury	-3.16	0.068	0.44	0.086	0.54	0.285	0.02	0.253	-0.45	0.253	-0.11	0.232	-3.85	0.240	0.953	572.9

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Reg. util.	SE	R2adj	AIC
Newcastle upon Tyne	-0.96	0.037	0.98	0.041	1.20	0.181	0.19	0.152	-1.20	0.137	-0.25	0.124	0.21	0.191	0.871	460.5
Newport	-1.51	0.102	0.67	0.103	1.81	0.445	-0.57	0.304	-0.43	0.390	0.43	0.317	-0.17	0.427	0.948	336.0
Northampton	-1.79	0.104	0.59	0.070	0.99	0.279	-0.75	0.199	-0.07	0.227	0.07	0.207	-1.03	0.303	0.839	413.0
Norwich	-3.22	0.253	0.78	0.142	1.01	0.461	-1.30	0.332	-0.87	0.420	-0.98	0.389	-1.06	0.615	0.751	399.1
Nottingham	-1.17	0.073	0.77	0.049	1.48	0.196	-0.09	0.165	-0.31	0.170	0.53	0.160	0.55	0.246	0.813	467.3
Oldham	-1.95	0.046	0.40	0.083	1.12	0.307	-0.89	0.196	-0.47	0.255	0.71	0.270	-2.72	0.244	0.969	504.7
Oxford	-0.53	0.081	0.80	0.053	2.83	0.197	-0.49	0.151	-0.85	0.138	0.80	0.136	-0.48	0.220	0.698	521.2
Peterborough	-1.05	0.182	0.56	0.085	0.24	0.300	-0.56	0.217	-0.71	0.251	-0.50	0.236	0.69	0.387	0.582	324.2
Plymouth	-1.13	0.066	0.80	0.058	-0.29	0.194	0.91	0.146	-0.56	0.192	-0.29	0.156	-3.14	0.422	0.671	478.3
Poole	-1.30	0.071	0.80	0.146	0.90	0.508	-1.18	0.391	-1.34	0.437	-1.20	0.366	-0.36	0.335	0.972	151.4
Portsmouth	-0.80	0.078	0.81	0.069	0.07	0.222	-0.73	0.218	-0.14	0.193	-0.25	0.173	1.02	0.269	0.494	482.8
Preston	-1.54	0.071	0.60	0.080	0.22	0.284	-0.22	0.216	-0.39	0.263	-0.09	0.238	2.17	0.324	0.782	423.6
Reading	-1.45	0.035	1.10	0.068	0.02	0.198	-1.61	0.190	-2.31	0.140	-2.65	0.121	0.03	0.172	0.916	755.7
Rochdale	-1.78	0.044	0.47	0.078	0.71	0.282	-0.95	0.197	-1.41	0.245	-0.36	0.236	-1.86	0.253	0.928	563.8
Rochester upon Medway	-2.25	0.073	0.17	0.104	1.11	0.232	-1.82	0.237	1.04	0.255	0.42	0.236	-1.64	0.264	0.933	471.1
Rotherham	-2.38	0.061	0.41	0.071	-0.77	0.297	-0.04	0.251	1.45	0.258	0.70	0.224	3.03	0.370	0.985	508.1
Salford	-1.83	0.036	0.91	0.062	-2.17	0.218	0.98	0.157	-0.31	0.183	-0.75	0.171	-2.12	0.174	0.891	724.0
Scarborough	-2.21	0.108	1.00	0.067	-0.88	0.287	0.87	0.220	-0.67	0.296	-1.20	0.205	-0.29	0.264	0.886	360.5
Sheffield	-1.75	0.027	0.58	0.038	1.09	0.145	-0.65	0.121	-0.24	0.127	0.50	0.117	2.31	0.177	0.919	744.4
Southampton	-1.24	0.082	0.62	0.072	1.58	0.235	-1.05	0.208	-0.09	0.204	0.50	0.175	0.19	0.246	0.793	390.6
Southwark	-1.13	0.025	0.91	0.055	1.20	0.128	-0.52	0.132	0.30	0.132	0.56	0.126	2.21	0.309	0.960	634.1
St Albans	-2.13	0.090	0.01	0.085	0.57	0.222	-1.62	0.210	0.90	0.206	0.24	0.207	-2.70	0.379	0.782	568.1
Stafford	-1.82	0.098	0.95	0.075	-1.81	0.270	-0.17	0.231	-1.09	0.270	-1.96	0.193	-1.94	0.285	0.906	487.9
Stirling	-1.31	0.095	1.25	0.103	2.61	0.553	-1.23	0.396	-0.84	0.311	-0.81	0.237	2.63	0.575	0.978	219.5
Stockport	-1.87	0.061	0.96	0.061	2.09	0.257	-0.38	0.147	-1.52	0.176	-0.59	0.171	-1.98	0.233	0.983	573.8
Stoke-on-Trent	-2.79	0.093	0.61	0.086	0.93	0.318	-2.00	0.195	-0.82	0.287	-0.79	0.265	-4.43	0.284	0.945	435.2
Stratford-on-Avon	-2.74	0.107	0.73	0.076	1.03	0.387	-0.52	0.246	-0.93	0.301	-0.60	0.262	-3.40	0.396	0.981	375.4
Sunderland	-1.34	0.051	0.85	0.060	0.02	0.263	0.06	0.196	-0.55	0.209	-0.58	0.182	1.56	0.220	0.920	407.4
Swansea	-1.77	0.122	0.86	0.077	1.57	0.328	0.24	0.193	0.05	0.292	0.54	0.263	0.44	0.486	0.913	315.3
Thamesdown	-1.55	0.096	0.64	0.074	0.49	0.289	-0.65	0.233	-0.49	0.238	-0.53	0.193	-0.09	0.261	0.666	433.9
Tower Hamlets	-0.79	0.046	0.48	0.072	-0.90	0.157	0.72	0.179	-0.43	0.167	0.16	0.158	1.00	0.425	0.909	555.4
Trafford	-1.65	0.047	0.60	0.061	2.88	0.217	-1.26	0.145	-1.84	0.184	-0.16	0.191	-1.52	0.245	0.982	584.3
Wakefield	-2.20	0.056	0.79	0.047	-0.43	0.244	-0.33	0.177	-0.31	0.210	-0.85	0.169	-2.85	0.199	0.958	595.9
Walsall	-1.69	0.045	0.71	0.051	1.06	0.361	-0.10	0.249	0.09	0.335	1.20	0.250	-0.03	0.490	0.985	487.8

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Reg. util.	SE	R2adj	AIC
Warrington	-1.90	0.062	0.72	0.079	0.48	0.271	-0.69	0.202	-1.36	0.242	-1.10	0.212	-2.30	0.239	0.902	482.5
Warwick	-2.43	0.074	0.52	0.071	1.91	0.352	-1.05	0.200	-0.41	0.229	-0.36	0.190	-3.22	0.341	0.987	392.9
Wigan	-2.10	0.057	0.67	0.075	-0.63	0.264	0.02	0.184	-1.33	0.230	-0.96	0.211	-2.61	0.215	0.928	519.3
Wokingham	-1.97	0.032	0.72	0.075	1.34	0.272	-2.21	0.224	-0.83	0.190	-1.40	0.151	-1.25	0.174	0.983	519.5
Wolverhampton	-1.97	0.043	0.40	0.059	0.55	0.249	-0.50	0.160	-0.07	0.245	0.34	0.216	-2.21	0.209	0.993	395.9
York	-1.57	0.085	0.59	0.060	2.11	0.257	-0.08	0.179	-1.12	0.219	0.35	0.187	0.75	0.297	0.857	440.3

Appendix H: Hybrid weighted nested logit model goodness-of-fit and parameter estimates values

This appendix tabulates in full the AIC and R^2_{adj} goodness-of-fit statistics and the parameter estimates (with their standard errors) arising from global and origin-specific calibrations of the weighted nested logit migration destination choice model for all migrants aged 16+ years.

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Access.	SE	Reg util.	SE	R2adj	AIC
Global	-1.30	0.002	0.88	0.005	0.90	0.019	-0.15	0.014	-0.41	0.016	-0.08	0.015	-0.41	0.005	0.07	0.023	0.830	91475.9
Aberdeen City	-1.55	0.120	1.26	0.061	3.21	0.314	0.70	0.309	-0.53	0.320	1.08	0.262	-0.66	0.071	0.21	0.317	0.961	408.5
Barking and Dagenham	-1.09	0.127	0.84	0.133	1.26	0.383	-2.42	0.344	-1.35	0.241	-0.77	0.273	-0.63	0.102	1.87	0.523	0.602	363.6
Barnsley	-2.79	0.079	0.36	0.073	-1.01	0.282	0.74	0.281	1.01	0.290	0.58	0.193	-1.07	0.078	3.78	0.312	0.950	485.0
Bath	-0.96	0.061	1.17	0.080	1.38	0.288	1.07	0.224	-1.25	0.231	0.28	0.200	-0.67	0.065	0.29	0.348	0.941	331.0
Birmingham	-1.33	0.026	1.12	0.041	0.54	0.137	0.81	0.127	-1.34	0.114	-0.31	0.091	-0.19	0.034	-0.96	0.136	0.878	754.3
Blackburn	-1.64	0.085	0.88	0.096	-1.14	0.353	1.79	0.302	0.54	0.351	0.57	0.283	-0.66	0.086	1.62	0.357	0.658	447.2
Blackpool	-1.26	0.064	0.50	0.072	0.16	0.265	-0.04	0.216	0.21	0.276	0.14	0.217	-0.05	0.060	0.39	0.362	0.727	449.2
Bolton	-1.98	0.044	0.71	0.064	-1.39	0.227	0.94	0.177	-0.20	0.226	-0.38	0.185	-0.50	0.070	-2.22	0.180	0.970	500.7
Bournemouth	-0.72	0.040	0.97	0.077	0.37	0.261	0.72	0.248	-0.51	0.224	-0.04	0.200	-0.30	0.062	-0.14	0.233	0.845	341.5
Bradford	-1.91	0.058	1.02	0.043	-0.27	0.183	1.14	0.143	-1.22	0.156	-0.19	0.136	-0.97	0.045	1.76	0.202	0.988	524.4
Brighton	-1.07	0.165	1.11	0.070	1.99	0.245	-0.07	0.235	-0.86	0.191	0.77	0.180	-0.49	0.069	0.33	0.365	0.768	391.1
Bristol	-0.81	0.048	0.98	0.048	1.09	0.167	0.83	0.119	-1.02	0.141	0.40	0.127	-0.44	0.037	-0.19	0.213	0.711	539.4
Bury	-1.93	0.044	0.48	0.073	1.95	0.284	-0.36	0.226	-2.08	0.227	0.38	0.234	-0.86	0.083	2.24	0.330	0.979	474.7
Cambridge	-1.47	0.154	0.92	0.058	1.86	0.218	1.43	0.195	-1.33	0.157	1.07	0.154	-0.87	0.070	2.55	0.351	0.692	471.9
Camden	-1.03	0.039	0.86	0.054	2.50	0.140	-0.91	0.124	-0.86	0.116	0.71	0.116	-0.42	0.044	1.19	0.359	0.974	577.0
Canterbury	-1.79	0.065	0.74	0.084	1.63	0.263	-0.10	0.239	-0.19	0.215	1.10	0.200	-0.78	0.064	2.06	0.298	0.957	372.0
Cardiff	-1.20	0.055	0.75	0.050	2.46	0.223	-0.20	0.133	-0.22	0.192	1.22	0.167	-0.34	0.048	0.27	0.295	0.901	468.5
Carlisle	-1.71	0.163	0.66	0.084	0.51	0.394	0.57	0.310	-0.46	0.307	-0.06	0.279	-0.29	0.077	0.52	0.357	0.553	317.8
Chelmsford	-3.13	0.147	0.86	0.090	0.94	0.266	0.05	0.293	-1.54	0.215	-0.49	0.212	-1.15	0.083	-4.77	0.412	0.862	392.3
Cheltenham	-1.93	0.059	0.83	0.080	1.13	0.344	0.20	0.255	-1.24	0.291	0.06	0.236	-0.56	0.065	-1.50	0.268	0.978	278.6
Chester	-1.23	0.090	0.92	0.069	1.96	0.260	-0.02	0.226	-1.60	0.223	0.07	0.197	-0.39	0.065	0.81	0.267	0.769	366.7
Colchester	-1.83	0.108	0.76	0.078	1.41	0.249	-0.09	0.244	-1.14	0.197	-0.06	0.196	-0.49	0.061	-2.10	0.307	0.696	439.4
Coventry	-1.99	0.070	0.38	0.043	2.66	0.194	-1.25	0.156	-0.25	0.178	0.90	0.154	-0.50	0.045	2.52	0.319	0.933	568.2
Croydon	-1.56	0.041	1.08	0.058	2.01	0.163	-0.99	0.177	0.27	0.133	0.91	0.126	-0.67	0.047	-2.19	0.252	0.900	628.6
Darlington	-1.62	0.098	1.00	0.099	0.23	0.376	0.68	0.331	-0.81	0.309	-0.51	0.263	-0.39	0.089	1.42	0.307	0.780	265.1
Derby	-1.35	0.087	0.74	0.066	1.13	0.268	0.23	0.211	-0.76	0.244	0.70	0.215	-0.70	0.060	0.48	0.297	0.907	298.9

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Access.	SE	Reg util.	SE	R2adj	AIC
Doncaster	-1.94	0.050	0.34	0.051	-0.04	0.236	1.19	0.201	-0.29	0.235	0.58	0.186	-0.35	0.056	-1.10	0.181	0.895	548.2
Dover	-1.70	0.066	0.80	0.118	0.89	0.332	-0.17	0.289	-0.22	0.318	0.28	0.268	-0.76	0.100	2.36	0.571	0.961	333.6
Dudley	-1.78	0.058	0.30	0.058	0.53	0.290	-0.39	0.248	0.08	0.314	0.98	0.239	-0.54	0.069	1.33	0.395	0.973	450.1
Dundee City	-1.48	0.097	1.12	0.072	1.43	0.392	-0.05	0.355	-1.51	0.186	-0.10	0.175	-0.45	0.079	1.71	0.406	0.988	263.1
Durham	-1.39	0.053	0.89	0.075	1.39	0.252	0.84	0.231	-1.00	0.242	0.18	0.212	-0.38	0.066	0.84	0.222	0.946	299.2
Ealing	-1.47	0.046	0.75	0.051	2.61	0.149	-1.42	0.139	0.46	0.127	1.27	0.121	-0.40	0.043	2.14	0.312	0.852	690.3
Edinburgh City	-1.01	0.060	1.09	0.050	2.78	0.206	-0.02	0.177	-2.31	0.106	-0.10	0.104	-0.52	0.038	0.59	0.193	0.895	511.4
Exeter	-0.99	0.118	0.98	0.078	0.58	0.236	1.09	0.197	-1.07	0.235	0.20	0.208	-0.48	0.063	-0.63	0.386	0.724	412.8
Glasgow City	-0.97	0.149	1.32	0.124	1.55	0.657	-0.60	0.588	-1.48	0.345	-0.54	0.347	-0.59	0.118	2.54	0.847	0.888	163.2
Gloucester	-2.10	0.074	0.85	0.093	-0.44	0.381	0.40	0.279	-0.93	0.348	-0.45	0.284	-0.40	0.073	-2.02	0.324	0.976	293.1
Greenwich	-0.83	0.057	0.94	0.068	1.00	0.185	0.15	0.189	0.08	0.150	0.88	0.151	-0.68	0.052	-0.70	0.315	0.731	541.7
Guildford	-0.91	0.134	0.85	0.078	1.05	0.248	1.26	0.245	-0.70	0.199	0.52	0.180	-0.64	0.073	1.42	0.372	0.702	364.1
Hackney	-1.00	0.037	0.90	0.065	0.93	0.130	-0.11	0.121	-0.98	0.145	0.48	0.149	-0.53	0.051	1.02	0.414	0.985	513.2
Hammersmith and Fulham	-1.41	0.044	1.17	0.058	1.15	0.158	-0.07	0.169	0.44	0.127	0.88	0.144	-0.54	0.055	2.06	0.285	0.901	759.8
Harrogate	-2.31	0.074	0.70	0.044	-0.26	0.238	1.46	0.170	-1.13	0.203	-0.43	0.175	-0.95	0.058	1.72	0.254	0.969	440.6
Harrow	-1.24	0.065	1.02	0.074	1.86	0.212	-0.95	0.201	0.52	0.173	0.88	0.167	-0.52	0.064	-0.80	0.568	0.914	512.7
Ipswich	-1.68	0.135	0.88	0.122	1.61	0.404	-0.53	0.366	-1.59	0.285	-0.09	0.300	-0.58	0.089	-0.93	0.408	0.879	194.8
Islington	-1.00	0.038	0.95	0.059	1.94	0.174	-0.33	0.169	-0.90	0.126	0.73	0.132	-0.60	0.047	0.96	0.342	0.983	542.0
Kensington and Chelsea	-1.42	0.034	0.62	0.064	2.14	0.176	-0.63	0.151	-0.03	0.134	0.99	0.151	-0.41	0.059	3.73	0.338	0.979	640.7
Kings Lynn & W. Norfolk	-2.53	0.196	0.80	0.153	-1.19	0.491	2.22	0.469	-1.94	0.344	-0.60	0.375	-1.19	0.107	-2.02	0.499	0.762	358.3
Kingston upon Hull	-1.79	0.128	1.07	0.056	1.05	0.256	0.51	0.217	-0.81	0.209	0.42	0.189	-0.57	0.060	0.92	0.325	0.852	345.3
Lambeth	-1.44	0.022	1.38	0.044	2.58	0.133	-1.52	0.124	-0.32	0.101	0.52	0.105	-0.53	0.039	2.18	0.319	0.963	720.1
Lancaster	-1.51	0.084	0.86	0.063	0.37	0.233	0.57	0.191	-0.70	0.213	-0.14	0.184	-0.28	0.053	1.02	0.278	0.779	404.5
Leeds	-2.19	0.032	1.07	0.033	0.25	0.120	0.55	0.091	-1.82	0.089	-0.67	0.078	-0.79	0.032	2.62	0.126	0.943	781.4
Leicester	-1.44	0.125	0.95	0.046	1.68	0.204	-0.08	0.149	-0.14	0.167	0.98	0.155	-0.48	0.050	1.74	0.357	0.687	490.3
Lincoln	-1.90	0.214	0.98	0.092	1.52	0.364	-0.02	0.289	-0.16	0.323	0.60	0.292	-0.83	0.083	1.19	0.552	0.705	296.5
Liverpool	-1.50	0.061	1.08	0.047	1.20	0.174	0.22	0.150	-1.41	0.154	0.14	0.141	-0.56	0.046	1.53	0.202	0.829	501.2
Luton	-2.40	0.096	0.95	0.078	0.81	0.289	-0.33	0.256	-0.43	0.234	0.15	0.224	-0.72	0.076	-2.64	0.305	0.948	357.4
Macclesfield	-2.13	0.076	0.82	0.072	2.08	0.280	0.43	0.254	-2.32	0.229	-0.54	0.199	-1.10	0.089	2.02	0.349	0.978	431.3
Maidstone	-1.95	0.062	0.58	0.101	-0.42	0.249	0.95	0.295	0.75	0.233	0.12	0.213	-0.47	0.063	-3.37	0.268	0.976	362.7
Manchester	-1.75	0.028	0.68	0.033	1.90	0.129	-0.30	0.110	-0.39	0.118	0.64	0.103	-0.37	0.039	2.55	0.135	0.958	798.7
Middlesbrough	-1.28	0.085	0.96	0.076	1.22	0.297	0.42	0.238	-1.02	0.257	0.15	0.244	-0.40	0.068	0.40	0.263	0.681	358.3
Milton Keynes	-2.17	0.096	0.77	0.071	0.49	0.275	0.46	0.247	-0.65	0.210	0.09	0.196	-0.58	0.063	-2.64	0.299	0.872	387.5
Newbury	-3.14	0.068	0.83	0.086	0.67	0.289	1.13	0.276	-1.50	0.268	-0.20	0.227	-0.81	0.071	-4.65	0.251	0.943	593.7

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Access.	SE	Reg util.	SE	R2adj	AIC
Newcastle upon Tyne	-0.95	0.036	1.04	0.042	1.11	0.178	0.70	0.167	-1.16	0.134	0.01	0.126	-0.27	0.036	0.04	0.183	0.873	459.6
Newport	-1.50	0.082	0.81	0.107	1.32	0.420	-0.17	0.259	-0.65	0.367	0.28	0.323	-0.48	0.089	1.80	0.608	0.945	342.9
Northampton	-2.08	0.097	0.77	0.067	0.73	0.283	0.15	0.227	-0.34	0.226	0.35	0.207	-0.54	0.056	-2.07	0.313	0.927	336.0
Norwich	-3.24	0.255	1.00	0.140	1.17	0.472	-0.64	0.362	-1.12	0.404	-0.58	0.385	-0.56	0.114	-1.78	0.649	0.718	412.0
Nottingham	-1.51	0.072	0.89	0.048	1.88	0.197	0.01	0.157	-0.83	0.171	0.73	0.157	-0.60	0.045	1.07	0.236	0.915	389.9
Oldham	-2.34	0.058	0.69	0.086	0.84	0.325	0.63	0.274	-1.24	0.252	0.81	0.264	-1.30	0.099	3.13	0.305	0.982	449.2
Oxford	-1.18	0.092	0.90	0.051	2.84	0.196	0.40	0.167	-1.23	0.139	1.27	0.138	-0.63	0.051	1.44	0.260	0.746	505.1
Peterborough	-1.66	0.170	0.81	0.083	0.23	0.316	0.25	0.259	-0.94	0.246	-0.10	0.237	-0.60	0.066	-0.67	0.397	0.684	297.4
Plymouth	-0.99	0.082	0.96	0.061	-0.47	0.187	1.06	0.151	-0.62	0.185	-0.27	0.153	-0.42	0.048	-0.71	0.316	0.715	465.0
Poole	-1.28	0.071	0.85	0.146	0.78	0.509	-0.68	0.464	-1.35	0.426	-1.05	0.366	-0.23	0.114	-0.70	0.380	0.972	152.0
Portsmouth	-0.82	0.071	1.08	0.068	-0.26	0.220	0.81	0.240	-0.64	0.189	0.01	0.167	-0.67	0.053	0.07	0.293	0.685	436.7
Preston	-1.65	0.072	0.78	0.081	0.25	0.292	0.52	0.232	-1.11	0.271	-0.03	0.230	-0.55	0.069	1.46	0.302	0.841	393.4
Reading	-1.72	0.042	1.21	0.064	-0.40	0.206	0.18	0.227	-2.49	0.136	-1.86	0.126	-0.80	0.053	-1.48	0.209	0.912	761.1
Rochdale	-2.15	0.059	0.66	0.077	0.34	0.295	0.66	0.229	-1.91	0.242	0.05	0.231	-1.01	0.097	-1.52	0.215	0.948	532.7
Rochester upon Medway	-2.48	0.074	0.69	0.102	0.32	0.260	0.14	0.297	0.01	0.253	0.60	0.227	-0.92	0.074	-3.21	0.298	0.970	393.5
Rotherham	-2.33	0.056	0.45	0.071	-0.77	0.274	-0.49	0.235	1.06	0.273	0.47	0.217	-0.76	0.066	2.89	0.352	0.979	542.7
Salford	-1.66	0.033	0.85	0.060	-1.67	0.246	1.42	0.197	-1.24	0.202	-0.73	0.169	-0.77	0.071	0.64	0.238	0.904	712.1
Scarborough	-2.26	0.138	1.09	0.070	-0.70	0.280	1.00	0.220	-0.52	0.302	-1.07	0.220	-0.29	0.072	-0.28	0.306	0.875	371.1
Sheffield	-2.33	0.033	0.80	0.036	1.85	0.153	-0.34	0.106	-0.81	0.128	0.96	0.116	-0.98	0.040	2.97	0.135	0.972	640.0
Southampton	-1.21	0.076	0.86	0.072	1.34	0.231	0.24	0.232	-0.76	0.211	0.57	0.171	-0.65	0.060	0.89	0.317	0.849	360.1
Southwark	-1.22	0.028	1.15	0.057	1.13	0.135	-0.11	0.139	0.17	0.123	0.74	0.124	-0.62	0.046	0.89	0.266	0.964	625.3
St Albans	-2.46	0.076	0.56	0.083	0.50	0.247	-0.41	0.234	0.01	0.192	0.42	0.196	-0.90	0.063	-3.93	0.328	0.911	480.1
Stafford	-2.38	0.114	0.81	0.072	-0.62	0.310	-0.03	0.240	-1.55	0.274	-1.12	0.209	-0.75	0.073	-2.22	0.286	0.968	383.8
Stirling	-1.30	0.110	1.18	0.106	2.40	0.559	0.53	0.493	-0.92	0.293	0.17	0.278	-0.71	0.094	0.58	0.595	0.987	170.4
Stockport	-2.09	0.058	0.98	0.059	1.64	0.249	0.71	0.191	-1.14	0.187	0.16	0.176	-1.02	0.063	2.22	0.264	0.994	471.3
Stoke-on-Trent	-3.29	0.104	0.68	0.080	0.45	0.334	-0.92	0.219	-1.31	0.292	-0.33	0.267	-0.90	0.084	-4.47	0.282	0.972	369.7
Stratford-on-Avon	-3.05	0.117	0.78	0.073	1.00	0.383	-0.34	0.246	-1.37	0.303	-0.42	0.258	-0.53	0.070	-4.07	0.421	0.981	376.1
Sunderland	-1.37	0.052	0.93	0.063	-0.11	0.263	0.42	0.216	-0.58	0.205	-0.45	0.183	-0.21	0.054	1.36	0.200	0.921	407.6
Swansea	-1.66	0.136	0.97	0.078	1.69	0.319	0.44	0.188	-0.20	0.286	0.68	0.255	-0.40	0.060	0.84	0.477	0.934	288.5
Thamesdown	-1.62	0.088	0.87	0.073	0.20	0.288	0.36	0.253	-1.08	0.233	-0.43	0.182	-0.71	0.058	-0.85	0.264	0.778	394.3
Tower Hamlets	-0.84	0.050	0.56	0.076	-0.89	0.158	0.80	0.180	-0.45	0.164	0.20	0.158	-0.17	0.053	0.85	0.394	0.902	564.3
Trafford	-1.83	0.047	0.75	0.062	2.65	0.223	-0.30	0.183	-1.89	0.178	0.19	0.185	-0.87	0.068	1.81	0.284	0.988	545.1
Wakefield	-2.68	0.072	0.51	0.054	0.36	0.262	0.61	0.189	-0.18	0.228	0.47	0.199	-1.30	0.069	3.22	0.265	0.976	542.3
Walsall	-1.89	0.044	0.52	0.050	0.86	0.361	0.24	0.272	-0.85	0.342	0.92	0.249	-0.42	0.073	-1.25	0.362	0.993	407.6

Origin	Dist.	SE	Pop.	SE	SClass	SE	HPrice	SE	Tenure	SE	Unemp.	SE	Access.	SE	Reg util.	SE	R2adj	AIC
Warrington	-2.07	0.068	0.84	0.079	0.31	0.280	0.04	0.227	-1.89	0.245	-1.07	0.202	-0.64	0.076	-2.08	0.248	0.916	468.2
Warwick	-2.59	0.074	0.55	0.067	1.76	0.358	-0.38	0.217	-0.56	0.236	-0.01	0.185	-0.45	0.065	-3.69	0.337	0.990	374.7
Wigan	-2.17	0.059	0.75	0.075	-0.54	0.277	0.43	0.225	-2.05	0.241	-1.02	0.204	-0.62	0.073	-2.28	0.245	0.927	521.7
Wokingham	-1.95	0.031	1.01	0.072	1.29	0.277	-0.61	0.254	-1.23	0.188	-0.91	0.150	-0.84	0.056	-2.14	0.185	0.983	522.9
Wolverhampton	-2.06	0.063	0.36	0.061	0.49	0.249	-0.30	0.156	-0.22	0.250	0.36	0.215	-0.12	0.061	-2.36	0.218	0.992	407.4
York	-2.01	0.094	0.72	0.055	2.44	0.253	0.85	0.180	-1.12	0.217	1.18	0.192	-0.81	0.061	1.21	0.284	0.917	386.9