A longitudinal mixed logit model for estimation of push and pull effects of area characteristics in residential location choice

Fiona Steele[1], Elizabeth Washbrook[2], William J. Browne[2] and Christopher Charlton[2]

[1]Department of Statistics, London School of Economics

[2]Centre for Multilevel Modelling, University of Bristol

Abstract

We consider households' choice of neighbourhood using household panel data linked to neighbourhood data. A type of random effects multinomial logit model is developed to study neighbourhood choice over time, extending previous research that has been restricted to cross-sectional data. We show how longitudinal data on individuals' residential location allow separation of the influence of neighbourhood characteristics on the decision to stay in the current area ('push' effects) and on the choice of destination among movers ('pull' effects). The model is applied in an analysis of residential choice in England, with a focus on how effects of area deprivation and distance from the current area on location choice at year t depend on characteristics at $t-1$ and life course transitions between $t-1$ and t. Household random effects are included to allow for unobserved heterogeneity between households in their propensity to move, and in the importance placed on deprivation and distance.

1. INTRODUCTION

The sorting of different types of household into different types of neighbourhood has fundamental implications for many social outcomes. Patterns of residential mobility shape the spatial distribution of populations and the extent to which certain groups, such as ethnic minorities, immigrants and the economically disadvantaged, are geographically concentrated. Any study of neighbourhood effects – the causal effect of place on people – must tackle the issue of non-random selection into places of residence. Indeed, it can be argued that the selection process is more than a statistical nuisance in neighbourhood effects research, and is in fact integral to understanding the key issues of interest (Bergström and van Ham 2010).

The availability of geographical identifiers in many datasets, most notably in national censuses, means it is relatively easy to document where different types of household are located. It is far more challenging, however, to understand *why* households have chosen the neighbourhoods in which they live. The difficulty arises because neighbourhoods are multidimensional "packages" of different attributes, in which types of dwellings, physical geography and social composition are chosen simultaneously. The number of alternatives from which a household can realistically choose is limited, and trade-offs between different area characteristics when making a decision are inevitable. Disentangling which are the decisive factors in attracting a household to (or repelling it from) a particular neighbourhood is therefore not straightforward. The problem has long been recognised in sociological research on racial segregation in the US, where scholars have sought to discriminate between different explanations for the apparent aversion of whites to living in neighbourhoods with high proportions of non-white residents (South and Crowder 1997).

Much of the previous research on residential location choice has abstracted from the more complicated features of this decision process. One approach has focused on the decision to exit the current neighbourhood, effectively grouping all alternative destinations into a single category (Lee, Oropesa and Kanan 1994, Crowder and South 2008, van Ham and Clark 2009). Using the terminology of Lee, et al. (1994), this type of analysis can identify factors that "push" households to leave a neighbourhood, but is silent on what attracts or "pulls" them towards a particular alternative. This question can be addressed in a limited way by dividing the set of potential destinations into a small number of groups and modelling the probability of moves to particular types of neighbourhood, for example with different levels of immigrant concentration or poverty (e.g. South, et al. 1997, Quillian 2003, Bolt, van Kempen and van Ham 2008). A second approach is to focus on the sample of movers only, and model the difference in a particular attribute between the origin and destination neighbourhoods (Clark 1992, Clark and Ledwith 2007). A number of studies have combined analysis of the decision of whether to move with analysis of neighbourhood change following a move to gain a more rounded picture of the nature of push and pull factors on location choice. Such analyses are usually carried out in two separate stages (South and Crowder 1998, Rabe and Taylor 2010); joint models that allow for dependence between the decision to move and choice of destination are rare (Crowder, South and Chavez 2006).

The studies discussed above are limited by the fact that destination neighbourhoods are generally characterised along a single dimension. An approach that overcomes this limitation is to use a conditional logit model, a type of multinomial logit model where the categorical response is the chosen neighbourhood (rather than neighbourhood type) and neighbourhood characteristics are included as explanatory variables (Nechyba and Strauss 1998, Ioannides and Zabel 2008, Hedman, van Ham and Manley 2011, Bruch and Mare 2012). Conditional logit models are especially well suited to the study of residential location choice where households evaluate a set of neighbourhoods as prospective destinations on multiple dimensions, the importance of which may vary according to household characteristics. For example, Hedman, et al. (2011) examined the extent to which households choose (or are restricted to) neighbourhoods with similar characteristics to their own. However, previous applications of conditional logit models have either considered only cross-sectional data or, where longitudinal data are available, have not fully exploited information on repeated residential choices.

Households' residential preferences, and constraints on their ability to secure accommodation that meets their needs, vary over time in response to changes in their economic and demographic characteristics. At the same time, neighbourhoods change in their desirability and affordability. Thus cross-sectional data on residential location and household and neighbourhood characteristics do not allow inferences about predictors of residential location choice. Longitudinal data from household panel studies and population registers provide rich information on changes in households' place of residence and other characteristics, which can be used to investigate the complex relationship between housing transitions and changes in household circumstances. Furthermore, longitudinal data can be used to separate effects of covariates on the decision to stay in the current area (push effects) from effects on the decision to move to a new area (pull effects), and to allow for unobserved heterogeneity between households in the importance they place on different neighbourhood attributes in location decisions. More generally, the concept of push and pull effects may be useful in any setting where a subject can decide to remain with their previous choice or switch to another, for example in studies of brand loyalty in market research. While mixed logit models have been developed for the analysis of longitudinal discrete-choice data (Jain, Vilcassim and Chintagunta 1994, Bhat and Guo 2004), to our knowledge, the distinction between push and pull effects of alternative-specific attributes has not been considered previously, nor have these models been used to study residential location choice.

In this article we present a general longitudinal mixed logit model for residential location choice, and show how the model can be parameterised to provide estimates of push and pull effects of area characteristics and their interaction with household characteristics. The model includes household random effects to account for unobserved heterogeneity in the effects of neighbourhood characteristics which may lead to dependency in household choices over time. The random effects are assumed to follow a multivariate normal distribution which allows for correlation between push and pull effects of a given neighbourhood characteristic and between the effects of different dimensions of neighbourhood quality. We show that non-zero random effect correlations also have an important role to play in relaxing the 'independence of irrelevant alternatives' assumption.

One likely reason for the lack of use of mixed models in neighbourhood choice research is the computational challenges in fitting random effects models to large-scale datasets that arise from a combination of a large choice set, large sample size, and long observation period. We propose an efficient, flexible Bayesian estimation procedure and provide software for estimation of this and more general models to longitudinal data from panel studies or population registers with large choice sets. The model is illustrated in an analysis of residential choice in England between 1998 and 2008 using data from the British Household Panel Survey.

2. A LONGITUDINAL MIXED LOGIT MODEL FOR RESIDENTIAL LOCATION CHOICE

2.1 Preliminaries

The following models are described in terms of household rather than individual choices, while recognising that co-resident non-related adults may be independent decision makers with regard to residential mobility and neighbourhood choice. We return to this issue, and give our working definition of a household, when we consider the application in Section 3.

Suppose that household i ($i = 1, \dots n$) chooses its area of residence between waves $t-1$ and t ($t = 2, \dots . T$) from a set of potential areas C_{it-1} containing R_{it-1} areas. The choice set is permitted to vary across households and time because it is both behaviourally unrealistic and computationally infeasible for households to choose from a common fixed set of areas (Lee and Waddell 2010). For example, the choice set might be restricted to the set of areas within a specified distance of a household's location at $t-1$.

Let y_{it} be the categorical response indicating the observed area of residence for household i at t. A general discrete choice model for the response probability is

$$\Pr(y_{it} = r) = p_{rit} = \frac{\exp(\eta_{rit})}{\sum_{k \in C_{it}} \exp(\eta_{kit})}, \qquad r = 1, \dots R_{it} \qquad (1)$$

where η_{rit} is the linear predictor which will usually be a function of area characteristics and their interactions with household characteristics.

The model in Equation (1) can also be expressed in terms of the log-ratio of the choice probabilities for a pair of areas r and s:

$$\log\left(\frac{p_{sit}}{p_{rit}}\right) = \eta_{sit} - \eta_{rit} \qquad (2)$$

2.2 A simple multinomial logit model with push and pull effects of area characteristics

We begin with a model that includes only area characteristics, but allow their effects on the choice between areas r and s at t in Equation (2) to depend on whether a household is resident in one of these areas at $t-1$. We show how this distinction, possible only with longitudinal data, allows estimation of "push" and "pull" effects of area characteristics.

Let $w_{rit-1} = I(y_{it-1} = r)$ where $I(.)$ is the indicator function, and denote by $\boldsymbol{z}_{rt-1}$ a p-vector of characteristics of area r defined at wave $t-1$. More generally, area characteristics can also be household specific, for example the distance between area r and the place of work for the head of household i. The linear predictor for the first model we consider is

$$\eta_{rit} = w_{rit-1}\,(\alpha + \boldsymbol{\beta}^T \boldsymbol{z}_{rt-1}) + (1 - w_{rit-1})\,\boldsymbol{\gamma}^T \boldsymbol{z}_{rt-1} \qquad (3)$$

where α is a scalar and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are parameter vectors. The type of model defined by Equations (1) and (3), with covariates relating to response alternatives and fixed coefficients across alternatives, was originally referred to as a conditional logit model (McFadden 1974). However, it was subsequently shown to be equivalent to the multinomial logit model, which traditionally has subject-specific covariates and alternative-specific coefficients (Maddala 1983). In common with most of the discrete-choice literature, we therefore refer to the model as a multinomial logit model hereafter. These models are widely used for analysing categorical responses where interest lies in the effects of attributes of the response alternatives on individual choice, including applications to brand preference (e.g. Jain, et al. 1994) and transportation demand (e.g. Ben-Akiva and Lerman 1985).

In the case of residential choice, Bruch, et al. (2012) first proposed the inclusion of the lagged choice indicator w_{rit-1} and its interaction with area characteristics $\boldsymbol{z}_{rt-1}$. We now show how a re-parameterisation of their model, given by Equation (3), allows separation of push and pull effects of $\boldsymbol{z}_{rt-1}$. A household's choice between two areas r and s at t depends on their residence at $t-1$ according to the values of w_{rit-1} and w_{sit-1} as follows.

Case 1: Resident in r at $t-1$ ($w_{rit-1} = 1,\ w_{sit-1} = 0$)

From Equations (2) and (3), the log-ratio of the probabilities of moving to area s between $t-1$ and t versus remaining in area r is

$$\log\left(\frac{p_{sit}}{p_{rit}}\right) = \boldsymbol{\gamma}^T \boldsymbol{z}_{st-1} - (\alpha + \boldsymbol{\beta}^T \boldsymbol{z}_{rt-1}) \qquad (4)$$

where $\boldsymbol{\gamma}$ may be interpreted as the effect of characteristics of potential area s on moving *to* that area (which we refer to as the "pull" effect of $\boldsymbol{z}$) and $-\boldsymbol{\beta}$ is the effect of characteristics of the current area r on moving *out of* that area ("push" effect of $\boldsymbol{z}$). α is the baseline log-probability ratio of staying in r rather than moving to a new area which we call the "inertia" parameter; the estimate of α is expected to be large and positive because most households do not change area between $t-1$ and t.

Case 2: Resident in neither r nor s at $t-1$ ($w_{rit-1}=0,\ w_{sit-1}=0$)

The log-ratio of the choice probabilities for area s versus area r at t, when resident in neither at $t-1$, is

$$\log\left(\frac{p_{sit}}{p_{rit}}\right)=\boldsymbol{\gamma}^T(\boldsymbol{z}_{st-1}-\boldsymbol{z}_{rt-1}) \qquad (5)$$

where, as in Case 1, $\boldsymbol{\gamma}$ may be interpreted as a pull effect, but now of one new area s over another r.

2.3 Allowing for household heterogeneity in effects of area characteristics: Differential push and pull effects

The linear predictor in Equation (3) can be extended to allow the push and pull effects of $\boldsymbol{z}_{rt-1}$ to depend on a q-vector of household characteristics $\boldsymbol{x}_{it-1}$:

$$\begin{aligned}\eta_{rit} &= w_{rit-1}\left\{\alpha+\boldsymbol{\beta}_0{}^T\boldsymbol{z}_{rt-1}+\boldsymbol{\beta}_1{}^T(\boldsymbol{x}_{it-1}*\boldsymbol{z}_{rt-1})\right\}\\ &\quad+(1-w_{rit-1})\left\{\boldsymbol{\gamma}_0{}^T\boldsymbol{z}_{rt-1}+\boldsymbol{\gamma}_1{}^T(\boldsymbol{x}_{it-1}*\boldsymbol{z}_{rt-1})\right\} \qquad (6)\end{aligned}$$

where $(\boldsymbol{x}_{it-1}*\boldsymbol{z}_{rt-1})^T=\left[x_{1it-1}\,\boldsymbol{z}_{rt-1}{}^T \quad \dots \quad x_{qit-1}\,\boldsymbol{z}_{rt-1}{}^T\right]$ is the qp-vector formed by taking the element-wise product of $\boldsymbol{x}$ and $\boldsymbol{z}$.

$\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are the push and pull effects of $\boldsymbol{z}$ when $\boldsymbol{x}=0$. Writing $\boldsymbol{\gamma}_1{}^T=\left[\boldsymbol{\gamma}_{11}{}^T \quad \dots \quad \boldsymbol{\gamma}_{1q}{}^T\right]$, $\boldsymbol{\beta}_0+\boldsymbol{\beta}_{1k}$ is the push effect of $\boldsymbol{z}$ for a 1-unit change in household characteristic x_k ($k=1,\dots,q$), and $\boldsymbol{\gamma}_0+\boldsymbol{\gamma}_{1k}$ is the corresponding pull effect.

It is straightforward to extend Equation (6) to include alternative-specific intercepts and coefficients of $\boldsymbol{x}$. We do not consider such models here because in the application to residential location choice their addition will be impractical due to the size of the choice set. Without fixed coefficients, however, it is not possible to estimate main effects of $\boldsymbol{x}$ because these cancel in the contrast between areas s and r in Equation (2).

2.4 The Independence of Irrelevant Alternatives assumption

It follows from Equations (1) and (2) that the log ratio of the choice probabilities for areas s and r depends on measured characteristics of these areas, but not of other potential areas. This property of the discrete-choice model is known as the "independence of irrelevant alternatives" (IIA), and it implies that the choice between s and r is unaffected by the addition or exclusion of other alternatives (Ben-Akiva, et al. 1985).

The source of the IIA assumption can be seen more clearly if the model is expressed in terms of continuous latent choice propensities (or utilities) y^*_{rit} which underlie the observed choice y_{it} such that $y_{it} = r$ if $y^*_{rit} > y^*_{sit}$ for $r \neq s$. The discrete choice model given by Equation (1) can be written

$$y^*_{rit} = \eta_{rit} + \epsilon_{rit}$$

where ϵ_{rit} are i.i.d. residuals, assumed to follow a Type I extreme value distribution with variance $\pi^2/6$. The IIA property stems from the independence of ϵ_{rit}, and will be invalid if the latent propensities to choose areas r and s are correlated. In the context of residential choice, non-zero residual correlation may arise because of similarity between areas on unmeasured factors used by households in deciding where to live.

Various approaches have been proposed to relax the IIA assumption, including generalised extreme value (GEV) models and the multinomial probit model. The nested logit model is the most widely used type of GEV model (e.g. Hensher, Rose and Greene 2005), but it requires the researcher to *a priori* partition subsets (nests) of similar alternatives within which IIA might reasonably hold. While the nested logit model is of limited use in modelling residential choice because of the difficulty in identifying areas that might be similar on unmeasured characteristics, it has been used to model residential location choice jointly with related decisions such as mode of transport and time of travel for various type of activity (Ben-Akiva and Bowman 1998) and residential mobility (Lee, et al. 2010). Both applications embed a multinomial logit model for residential choice within a nested logit model for the joint decision processes. The multinomial probit model allows explicitly for residual correlation by assuming that ϵ_{rit} follow a multivariate normal distribution, but this is infeasible in situations where the choice set is large. A more flexible way of accommodating similarity between alternatives is to use a mixed model which allows for unobserved heterogeneity between households in the effects of area characteristics.

2.5 Unobserved household heterogeneity

The most popular models for unobserved between-subject heterogeneity are the normal-mixed model and the latent class model (see Hensher and Greene (2003) and Keane and Wasi (2012a) for reviews of mixed logit models). The normal-mixed model includes normally distributed random coefficients for the effects of $\boldsymbol{z}_{rt-1}$. Log-normal distributions may be assumed for coefficients that are expected to have the same sign for all subjects. The latent class model avoids parametric distributional assumptions and assumes that subjects come from a finite set of subpopulations (Greene and Hensher 2003, Domanski and von Haefen 2010). We focus on normal-mixed models for several reasons. First, a large number of latent classes may be required to capture complex patterns of heterogeneity. In the application to residential choice, for example, there may be between-household differences in push and pull effects of multiple area characteristics, leading to a multidimensional finite mixture distribution. Second, the estimation and interpretation of latent class models is complicated by having a separate set of parameters for each class. Third, the direction and magnitude of the correlations between multivariate normal random effects are of direct substantive interest. For instance, correlated random effects can provide insights into the nature of the association between the household-specific push and pull effects of an area attribute z, or between the push or pull effects of two different area attributes.

While mixed logit models have been applied to cross-sectional data, identification of unobserved household heterogeneity will generally be much improved by the availability of longitudinal data. (See Revelt and Train (1998) for discussion and applications of mixed logit models with normal and log-normal random effects for repeated choice data.) In a normal-mixed model the coefficients of $\boldsymbol{z}$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ in Equation (6), are replaced by

$$\boldsymbol{\beta}_{0i} = \boldsymbol{\beta}_0 + \boldsymbol{u}_{\beta i}$$

$$\boldsymbol{\gamma}_{0i} = \boldsymbol{\gamma}_0 + \boldsymbol{u}_{\gamma i}$$

where $\boldsymbol{u}_{\beta i}$ and $\boldsymbol{u}_{\gamma i}$ are vectors of household-specific random effects which capture variation between households in the importance placed on $\boldsymbol{z}$ in location decisions. We also allow for unobserved heterogeneity in households' attachment to their current areas by replacing the inertia parameter α with the random coefficient

$$\alpha_i = \alpha + u_{\alpha i}$$

The random effects $\boldsymbol{u}_i{}^T = \begin{bmatrix} u_{\alpha i} & \boldsymbol{u}_{\beta i}{}^T & \boldsymbol{u}_{\gamma i}{}^T \end{bmatrix}$ are assumed to follow a multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{\Omega}_u$. Thus the linear predictor can be partitioned as $\eta_{rit} = \mu_{rit} + \delta_{rit}$ where μ_{rit} is the systematic (or "fixed") component given by Equation (6) and δ_{rit} is a random component which varies across households:

$$\delta_{rit} = w_{rit-1}\left(u_{\alpha i} + \boldsymbol{u}_{\beta i}{}^T \boldsymbol{z}_{rt-1}\right) + (1 - w_{rit-1})\ \boldsymbol{u}_{\gamma i}{}^T \boldsymbol{z}_{rt-1} \qquad (7)$$

From (7), δ_{rit} also varies over time, but only through the observed predictors w_{rit-1} and $\boldsymbol{z}_{rt-1}$.

We now show how the inclusion of household random effects relaxes the IIA assumption by considering how they affect the ratio of (conditional and marginal) choice probabilities and the correlation between latent choice propensities.

Ratio of choice probabilities. Equation (1) with η_{rit} defined by the sum of (6) and (7) gives the probability of choosing area r conditional on the household random effects $\boldsymbol{u}_i$. The addition of random coefficients $\boldsymbol{u}_{\beta i}$ and $\boldsymbol{u}_{\gamma i}$ allows the ratio of the subject-specific choice probabilities for areas r and s to vary across households according to differences in the (unobserved) importance placed on observed area characteristics $\boldsymbol{z}$. The IIA property is still assumed to hold at the household level because the ratio of subject-specific probabilities for areas r and s does not depend on characteristics of any other area. However, this is not the case for the ratio of unconditional or *marginal* choice probabilities. Letting $\eta_{rti}(\boldsymbol{u}_i)$ denote the linear predictor for the random effects model, the marginal (population-averaged) response probability is given by

$$\Pr(y_{it} = r) = \int \frac{\exp[\eta_{rit}(\boldsymbol{u}_i)]}{\sum_{k \in C_{it}} \exp[\eta_{kit}(\boldsymbol{u}_i)]} . \phi(\boldsymbol{u}_i)\, d\boldsymbol{u}_i \qquad (8)$$

where $\phi(.)$ is the standard normal pdf.

The log-ratio of the marginal response probabilities for areas r and s is no longer simply the difference in the linear predictors, as in Equation (2), because the summation in the denominator of (8) does not cancel. Thus the ratio of the marginal probabilities for r and s will depend on characteristics of other areas, and the IIA assumption is relaxed at the population level (Train 2003).

Correlation between latent choice propensities. The inclusion of household-specific effects induces a correlation between the latent choice propensities for any pair of areas r and s because $\boldsymbol{u}_i$ is common across the response alternatives faced by household i at time t. Consider a simplified form of Equations (6) and (7) with one area characteristic z_{rt-1}, leading to three random effects $(u_{\alpha i},\ u_{\beta i},\ u_{\gamma i})$ with covariance matrix

$$\Omega_u = \begin{pmatrix} \sigma_\alpha^2 & & \\ \sigma_{\alpha\beta} & \sigma_\beta^2 & \\ \sigma_{\alpha\gamma} & \sigma_{\beta\gamma} & \sigma_\gamma^2 \end{pmatrix}$$

Random effect covariances have received little attention in previous applications of mixed logit models and, indeed, random effects are commonly assumed to be independent. A notable exception is Revelt, et al. (1998) who contend that correlation between random effects would generally be expected. In the present application, the random effect covariances are of particular interest because they provide information about the associations between households' latent mobility preferences and the importance they place on z in residential decisions. For example $\sigma_{\beta\gamma} > 0$ implies a positive association between the push and pull effects of z, adjusting for the moderating effects of $\boldsymbol{x}$. The random effect covariances also play a crucial role in relaxing the IIA assumption, as shown below.

The covariance between the propensities to choose areas r and s for household i at t depends on a household's residence at $t-1$ as follows.

Case 1: Resident in r at $t-1$ $(w_{rit-1} = 1,\ w_{sit-1} = 0)$

For a household considering a move from area r to s between $t-1$ and t, the covariance between the latent propensities of remaining in r and moving to s is

$$\text{cov}(y_{rit}^*, y_{sit}^*) = \text{cov}\big(u_{\alpha i} + z_{rt-1}u_{\beta i}, z_{st-1}u_{\gamma i}\big) = z_{st-1}\sigma_{\alpha\gamma} + z_{rt-1}z_{st-1}\sigma_{\beta\gamma} \qquad (9)$$

when $\text{cov}(\epsilon_{rit}, \epsilon_{sit}) = 0$ and $\text{cov}(\delta_{rit}, \epsilon_{sit}) = 0$ for all $r \neq s$.

Thus the covariance depends on two components: (i) the value of z in the potential area, weighted by the covariance between the household-specific mobility propensity $(u_{\alpha i})$ and importance of z as a pull factor $(u_{\gamma i})$, and (ii) the similarity between areas r and s on z, weighted by the covariance between household-specific importance of z as a push and pull factor. If z is mean centred, the second component of the covariance will be highest for two areas with extreme above-average or below-average values on z and zero for two average areas.

Case 2: Resident in neither r nor s at $t-1$ $(w_{rit-1} = 0,\ w_{sit-1} = 0)$

The covariance between the latent propensities of choosing between two potential areas when currently resident in neither is

$$\text{cov}(y_{rit}^*, y_{sit}^*) = \text{cov}\big(z_{rt-1}u_{\gamma i}, z_{st-1}u_{\gamma i}\big) = z_{rt-1}z_{st-1}\sigma_\gamma^2 \qquad (10)$$

Thus the covariance between the latent attractiveness of two potential areas depends on their similarity with respect to z and on the between-household variance in the pull effect of z.

2.6 Unobserved area heterogeneity

The model may be further extended to allow for the effects of unmeasured neighbourhood characteristics on location choice by including choice-specific random effects $v_r \sim N(0, \sigma_v^2)$ in the linear predictor. However, it can be seen from the expressions for the covariance between the latent propensities of choosing areas r and s given by Equations (9) and (10) that the inclusion of area-specific random effects does not help to relax the IIA assumption unless $\text{cov}(v_r, v_s) \neq 0$. A natural extension would be to impose a spatial autocorrelation structure on the area effects, for example allowing for a non-zero covariance between neighbouring areas. Such spatial correlation would arise if areas in close proximity share unmeasured attributes that influence a household's location choice. Bhat, et al. (2004) proposed a mixed spatially correlated logit model for residential choice at a cross-section that includes a dissimilarity parameter measuring the correlation between adjacent areas.

One issue when considering the addition of choice-specific effects in applications where the choice set is large is that the number of choices can exceed the number of decision-makers, leading to an identification problem because many potential areas were not chosen by the survey respondents over the observation period. This is the case in our application where there are just over 6000 households and over 30000 areas. It is for this reason that we do not pursue the inclusion of area random effects, although details of estimating a more general model with (independent) area effects are given in the Appendix.

2.7 Sampling alternatives in large choice sets

Estimation of a multinomial logit model with alternative-specific attributes requires the data to be structured so that there is a record for each of the R_{it} response alternatives for household i at wave t. This can lead to a prohibitively large analysis file when the choice set is large, especially when decisions are observed over a long period for a large sample of households. A useful consequence of the IIA property, however, is that consistent parameter estimates can be obtained from a random subset of the full choice set, selected without replacement and including the record corresponding to the chosen alternative (McFadden 1978). For each household-wave it, denote by q_{rit} the probability that the record for alternative r is selected, where $q_{rit} = 1$ if $y_{it} = r$. McFadden described a situation where the choice set is fixed and $q_{rit} = q$. More generally, we may wish to include information about the likelihood that household i chooses alternative r at t, referred to as importance sampling (Ben-Akiva, et al. 1998, Bhat, Govindarajan and Pulugurta 1998, Brownstone, Bunch and Train 2000). In residential location choice, for example, the choice set for most households is restricted to areas within commuting distance of the residents' current place of work. This leads to substantial variation in R_{it} across households and time, where R_{it} will typically be considerably larger in metropolitan areas than in rural areas. In such cases q_{rit} will be proportional to R_{it}, and unequal selection probabilities are accommodated in the model by including $-\ln(q_{rit})$ as an offset term (e.g. Ben-Akiva, et al. 1985, Bruch, et al. 2012).

Unfortunately when the IIA assumption is relaxed, for example by including household-specific random coefficients, McFadden's theoretical result no longer holds and random sampling of choice sets may yield inconsistent estimates (Nerella and Bhat 2004, Keane and Wasi 2012b). This has led authors to explore empirically the impact of the size of the sampling fraction q on parameter estimates and standard errors from mixed multinomial logit models. Nerella, et al. (2004) conducted a simulation study with a cross-section of 750 individuals, a choice set of size 200 and q varying between 2.5% and 75%. They found a substantial impact of q on the bias and efficiency of the estimated parameters, and suggested that q should be set at 25% as a minimum. However, other research suggests that reliable estimates may be obtained using much smaller sampling fractions. In an analysis of vehicle choice with a total set of 689 models and makes, Brownstone, et al. (2000) used a random subset of 28 (4%) and reported that increasing the sampling fraction had little effect on parameter estimates, although it was important to stratify by vintage (one of the attributes of interest) because of a small number of new cars in the choice set. Keane, et al. (2012a) considered the impact of using random subsets of the choice set for three alternative mixed MNL models for panel data, including the normal mixed model, through Monte Carlo simulation and sensitivity analysis of real data. Based on simulations with 200 individuals, 20 choice occasions and 60 alternatives, biases were small when random subsets of 10 or 20 alternatives were used.

2.8 MCMC estimation

The most commonly used approaches for fitting mixed logit models are maximum simulated likelihood and MCMC estimation. Train (2001) compares these approaches and favours MCMC for both theoretical and computational speed reasons. He gives an MCMC algorithm for such models which generalizes work by Allenby (1997), and builds on ideas of Albert and Chibb (1993). In this paper, we modify Train's algorithm to accommodate parameterisations designed to improve the efficiency of MCMC estimation, which is especially important when the sample size and choice set are large. We consider a combination of hierarchical centering and orthogonal parameterisation, adapting algorithms used for estimation of multilevel binary-response models (see, for example, Browne, Steele, Golalizadeh and Green 2009).

The linear predictor given by Equation (6) includes parameters for inertia (α), and push effects ($\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$) and pull effects ($\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$) of choice-specific attributes. From an algorithmic point of view, it is convenient to distinguish between coefficients that have an associated individual-specific random effect ($\alpha, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0$), as described in Section 2.5, and those with a fixed effect only ($\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1$). Train (2003) focuses on a general model where all coefficients are random at the individual level, but considers the above specification as a special case that may be useful in certain situations, such as when the full random effects covariance matrix cannot be identified. In the application that follows, the variances of the random effects for $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$, the main push and pull effects of area attributes $\boldsymbol{z}_{rt-1}$, are of particular interest as measures of between-household heterogeneity in the effects of $\boldsymbol{z}_{rt-1}$ that are unexplained by household covariates $\boldsymbol{x}_{it-1}$.

Let $\boldsymbol{\theta}^T = \begin{bmatrix}\alpha & \boldsymbol{\beta}_0{}^T & \boldsymbol{\gamma}_0{}^T\end{bmatrix}$ and $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{u}_i$ with associated data vector $\boldsymbol{A}_{rit}$, and let $\boldsymbol{\varphi}^T = [\boldsymbol{\beta}_1{}^T \quad \boldsymbol{\gamma}_1{}^T]$ with data vector $\boldsymbol{B}_{rit}$. The linear predictor for the mixed logit model can be re-expressed as

$$\eta_{rit} = \boldsymbol{\theta}_i^T \boldsymbol{A}_{rit} + \boldsymbol{\varphi}^T \boldsymbol{B}_{rit} - \ln(q_{rit})$$

where $\boldsymbol{\theta}_i \sim MVN(\boldsymbol{\theta}, \boldsymbol{\Omega}_u)$ and $\ln(q_{rit})$ is an offset (see Section 2.7).

It is common to parameterise the model in terms of $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}$ rather than $\boldsymbol{u}_i$ and $\boldsymbol{\theta}$. This parameterisation is known as hierarchical centering (Gelfand, Sahu and Carlin 1995) and can improve mixing when the random effect variances Ω_u are not too small, as it allows a Gibbs sampling step for $\boldsymbol{\theta}$ rather than a Metropolis step. The other speed up we consider is an orthogonal reparameterisation similar to that in Browne, et al. (2009). This involves replacing $\boldsymbol{B}_{rit}$ by an orthogonal vector that spans the same space as $\boldsymbol{B}_{rit}$. This is achieved using a standard orthogonalising algorithm and we can then run MCMC using the transformed predictors. The chains for the parameters in the original parameterisation can be retrieved by a simple matrix transform based on the inverse of the transformation of the predictors (see Browne, et al. (2009) for details).

Full details of the MCMC algorithm are given in the Appendix. The algorithm has been implemented in the Stat-JR package (Charlton, et al. 2013) which allows MCMC chains to be run in parallel with both hierarchical centering and orthogonal parameterisation of the fixed predictors. The code has also been optimised so that estimation times are significantly faster than implementing the same model in an existing package such as WinBUGS (Lunn, Thomas, Best and Spiegelhalter 2000). Optimisation for the mixed logit model involves the storage of intermediate quantities and constants within the likelihood to reduce the number of computationally-expensive calculations, for example exponentiations. The likelihood also contains, for each observation, a linear predictor with many terms and several steps involve calculation of terms that are the linear predictor minus one element. Storage of the linear predictor for each observation and a technique of subtracting the relative element, updating it and then adding it back to the linear predictor leads to a substantial reduction in computing time.

3. BETWEEN-HOUSEHOLD HETEROGENEITY IN THE EFFECTS OF AREA DEPRIVATION AND DISTANCE ON RESIDENTIAL CHOICE IN ENGLAND

3.1 Data

Neighbourhood definition and variables

A neighbourhood is defined in our study as a Lower Super Output Area (LSOA), which are based on the Output Areas (OAs) created for the 2001 Census. OAs are the smallest standard units used for presenting official local statistical information in the UK and are designed to have similar population sizes and be as socially homogenous as possible based on tenure of household and dwelling type. An LSOA is an agglomeration of 4 to 6 contiguous OAs, and contains between 400 and 1200 households by definition, and an average of 1500 individuals. There were 32,482 LSOAs in England by the 2001 definition.

We characterize neighbourhoods along two dimensions: deprivation and distance from a person's current residence. The measure of deprivation used is the LSOA-level English Index of Multiple Deprivation (IMD), a weighted combination of seven domain indices which capture different aspects of deprivation within a neighbourhood, relating to income; employment; health and disability; education, skills and training; barriers to housing and geographical access to services; crime; and living environment. The income and employment domain indices carry more weight in the calculation of the overall IMD than the others, both because they are the most robust indicators of deprivation and because previous research indicates that these are the domains that are most likely to contribute to deprivation (Dibben, et al. 2007). The IMD is a cross-sectional index, but three versions have been constructed for 2004, 2007 and 2010. The score we allocate to a particular LSOA depends on the household survey year, as described below, and the resulting IMD measure is normalized using the English LSOA-level average and standard deviation from 2007.

The second neighbourhood measure considered is the distance in kilometres between a household's LSOA of current residence and each alternative LSOA. These are the straight-line distances between the population-weighted centroids of each area, calculated from the Ordnance Survey centroid grid references provided by the Office for National Statistics (ONS). An important difference between our distance and deprivation measures is that distance is household specific and, as a consequence, only its pull effect can be defined. If z_{sit-1} denotes the distance between area s and the current location of household i at $t-1$, $z_{rit-1}=0$ and therefore β vanishes from the expression for the log-ratio of the probabilities of moving to area s rather than remaining in r (Equation 4).

In order to define the choice set of neighbourhoods relevant to a household we use the household's current Travel-to-Work-Area (TTWA). TTWA is a labour market area definition, derived from 2001 Census information on home and work addresses, and used by the ONS to reflect areas where the bulk of the resident population also work within the same area. The criteria for defining TTWAs is that generally at least 75% of an area's resident workforce work in the area and at least 75% of the people who work in the area also live in the area. The area must also have a working population of at least 3,500. TTWA boundaries are non-overlapping and contiguous, and cover the whole of the UK. TTWAs do cross national boundaries, and of the 243 that cover the entire UK, 166 contain at least one LSOA in England. In our study a household's choice set of neighbourhoods includes all English LSOAs within the current TTWA. The mean number of LSOAs per TTWA is 196, but TTWAs are substantially larger in metropolitan areas such as London with 5467 LSOAs.

Household panel data

Data on the characteristics and residential locations of households are taken from the British Household Panel Survey (BHPS) (ISER 2009). The BHPS is a nationally representative sample of about 5500 private households recruited in 1991, containing approximately 10,000 adults who are interviewed annually. If anyone splits from their original household to form a new household, then all adult members of the new household are also interviewed. Children in original households are interviewed when they reach 16 years of age. The core questionnaire elicits information on topics such as household composition, housing tenure, employment and income at each annual interview. Our analysis uses information from waves 8–18, covering the period 1998–2008. Earlier waves are excluded because of lack of comparable area-level IMD data in this period.

One challenge in the analysis of household panel data is how to define a household longitudinally when its composition may change over time. The usual approach to this problem is to follow individuals, rather than attempt to track households, with analyses based on person-year observations. However, this is inappropriate for couple households where decisions are likely to be made jointly while partners are co-resident (Steele, Clarke and Washbrook 2013). For this reason, the following analysis is based on person-year observations with records from any adult observed at two adjacent years $t-1$ and t, but with couples contributing only one person-year record while they remain together. Using this approach, couples are regarded as a single decision-making unit, and any other individual is treated as an independent decision-maker. Thus an individual living with unrelated adults is treated the same as an individual living on their own. The sample is further restricted to individuals aged between 18 and 59 at t.

In the residential mobility literature, it is usual to distinguish local or short-distance moves from longer-distance moves because the two types of moves have very different determinants. By restricting the choice set to LSOAs within the TTWA of residence at $t-1$, we focus on local moves within a given labour or housing market which tend to be triggered by family events such as the arrival or departure of a child (Clark and Huang 2003). However, this raises the question of how to handle moves that cross a TTWA boundary. Among the 31,674 person-wave observations in our eligible sample we observed 4038 cross-LSOA moves between t and $t-1$ (12.7% mobility rate), of which 1203 (30%) cross a TTWA boundary. We distinguish those that are essentially local moves in the sense that the distance is no more than the longest within-TTWA move observed in the data (45 km). We retain these short-distance cross-TTWA moves, which account for 499 or 41% of all cross-TTWA moves, but drop the remaining 704 cases. The mean moving distance for the retained, short-distance cases is 21 km, while for the excluded long-distance cases it is 162 km. The choice set for the cross-TTWA movers remaining in the dataset consists of all LSOAs within the current TTWA of residence at $t-1$ (none of which were chosen), plus the single LSOA that was chosen at t from the new TTWA.

Household-level characteristics considered in the analysis are for the most part defined at $t-1$. We define seven categories of household type, distinguishing single males and females; single parents (of either sex) of a child under 16; and couples with a resident youngest child aged 0-4, 5-10, 11-15 and 16 or more. Housing tenure is categorized as owned (outright or with mortgage), private rented, social (council or Housing Association) rented, and living with family. This last group consists of people who are a relative (other than a spouse) of the household reference person (HRP). The HRP is the person legally or financially responsible for the accommodation. Ninety-five percent of individuals in the 'living with family' group are children living in the parental home, with siblings of the HRP making up the largest group among the remaining 5%. We measure the gross household income over the previous 12 months as the combined incomes of the two members of a couple, or the individual income of a single individual. We also include some indicators of life event transitions that occur between $t-1$ and t, i.e. contemporaneously with the choice of location at t. These include the birth of a child, a move into home ownership, a move out of home ownership, a move out of the family home into social or private rental, and all other tenure transitions (e.g. between private and social rented accommodation). Indicators of partnership formation and dissolution between $t-1$ and t were considered in preliminary analysis, but were not retained as the additional effects of these variables on location choice proved to be insignificant.

Sampling choice sets

The analysis sample consists of 30,970 person-wave observations from 6249 individuals (treating couples as a single 'individual' as described above). Expanding the data to include one record for each LSOA in a household's choice set results in a person-wave-LSOA dataset of over 29 million observations. LSOAs were randomly sampled from this expanded dataset with probability inversely proportional to the size of the TTWA, while always retaining the records for the LSOAs of residence at $t-1$ and t. Thus, for household i resident in TTWA j at $t-1$ the probability that LSOA r is selected from their choice set is

$$q_{rit-1|j} = \begin{cases} \dfrac{c}{\sqrt{R_{it-1|j}}} & \text{if } y_{it-1} \neq r \text{ and } y_{it} \neq r \\ 1 & \text{if } y_{it-1} = r \text{ or } y_{it} = r \end{cases}$$

$R_{it-1|j}$ is the number of areas in the choice set of household i at year $t-1$ given residence in TTWA j, and the constant c is chosen so that the number of records in the person-wave-LSOA file is approximately equal to a target of m_{tar} according to

$$c = \frac{m_{tar}}{\sum_j m_j \sqrt{R_j}}$$

where R_j is the total number of LSOAs in TTWA j and m_j is the total number of person-wave-LSOA records in TTWA j. The following results are based on an analysis file with $m_{tar} =$ 800,000. To assess sensitivity of estimates to random sampling of the choice set, the analysis was repeated for two different random subsets of person-wave-LSOA records, the first with the same value of m_{tar} and the second with $m_{tar} = 1{,}600{,}000$.

3.2 Results

As described above, we focus on the effects of two neighbourhood characteristics on location choice: area deprivation, measured by IMD, and distance from a household's current residence. We allow for both observed and unobserved heterogeneity in the push and pull effects of IMD and the pull effect of distance through their interactions with the household characteristics $\boldsymbol{x}_{it-1}$ described in the previous section and the inclusion of household-specific random effects.

The results presented below are based on five parallel chains of 100,000 MCMC iterations, each using a different starting value, with a burn-in of 10,000. Uniform priors were assumed for all parameters. Convergence was assessed by visual inspection of the five chains for each parameter. Increasing the chain length was found to have little impact on the posterior estimates. The posterior estimates were also insensitive to selection of different random samples of the choice set and doubling the sampling fraction. Hierarchical centering and orthogonal parameterisation, separately and in combination, were considered in an attempt to improve mixing. For the fixed parameters, orthogonal parameterisation led to substantial reductions in the effective sample size (ESS) (Kass, Carlin, Gelman and Neal 1998). However, hierarchical centering was found to have little impact on ESS of the random effect variances and covariances, most likely because in the case the variance estimates are small. We therefore present results from using orthogonal parameterisation.

Push and pull effects of deprivation and distance

Table 1 shows posterior estimates of the regression coefficients from the full mixed logit model (Model 4). The sign of the parameters representing the push effects of neighbourhood deprivation (β_0 and β_{1k}) have been reversed so that positive estimates indicate an aversion to deprivation in one's current neighbourhood. The first row of Table 1 shows push and pull effects for the reference values of the household characteristics $\boldsymbol{x}_{it-1}$ (β_0 , γ_{01} and γ_{02}). For this group of individuals we find that, as expected, a higher level of deprivation in a neighbourhood is associated with an increase in the probability of out-mobility and a decrease in the probability of being chosen as a destination by movers. Individuals are also less likely to move to neighbourhoods that are far from their current place of residence. The remaining rows of Table 1 show the additional push and pull effects for changes in the values of $\boldsymbol{x}_{it-1}$. For example, the push effect of deprivation depends on income according to $0.301 + 0.016 \ln(\text{income})$, and the effect for a single female is reduced by 0.101 compared to a childless couple. In the following discussion, we regard a 95% credible interval that does not include zero as evidence of a differential push or pull effect.

There is little impact of household income on the push effect of deprivation or on the pull effect of distance. However, higher income is associated with a stronger aversion to deprivation when choosing a new area, most likely because higher-income households are better able to act on preferences towards living in better-off areas. This result is consistent with research in the United States, Sweden and Britain that finds household income constrains movers' access to more advantaged neighbourhoods (Ioannides, et al. 2008, Hedman, et al. 2011, Clark, van Ham and Coulter 2013).

Previous research on Britain suggests that area deprivation exerts a push effect on mobility among couples but not singles (Rabe, et al. 2010) while singles are more likely than couples to move to less advantaged areas (Clark, et al. 2013). However, we find little evidence of differential push and pull effects of deprivation for singles and couples without children: differences between singles and couples only emerge when they have dependent children. Deprivation in the current area of residence has a weaker effect on the decision to move out for single parents than for other household types, and there is also a suggestion that deprivation has a weaker deterrent effect on choosing a new area among couples with older children (aged 16+) than for households with a younger child or no children. Distance is a more important factor in movers' choice of destination for single parents and couples with school-age children than for other household types, which may reflect stronger local ties and a reluctance to move far from current schools among these families. A birth between years $t-1$ and t strengthens the push effect of deprivation during the same period, but also weakens aversion to deprivation when choosing a new area to live. These apparently contradictory findings may be due to a desire to move out of a deprived area among some new parents while others compromise on neighbourhood quality in the search for an affordable family home.

The largest source of heterogeneity in the importance of deprivation and distance in neighbourhood choice is housing tenure and changes in tenure. The coefficients of housing tenure at $t-1$ are interpreted as contrasts in push and pull effects between non-homeowners and owners who do not change tenure within the next year. Estimates of the effects of neighbourhood characteristics for tenure changers can be obtained by summing values in the MCMC chains for the reference category β_0 and the relevant coefficients of dummy variables for tenure at $t-1$ and change in tenure between $t-1$ and t. This is illustrated in Figures 1 and 2 which show posterior estimates of the push and pull effects of deprivation for selected categories of housing tenure at $t-1$ and t. We find that the push effect of deprivation is weaker among private and social renters than for home-owners (Table 1 and Figure 1). Furthermore, private renters are less averse to deprivation and are prepared to move longer distances when choosing a new area (Table 1). In Britain private renters tend to be more mobile than owner-occupiers (e.g. Rabe, et al. 2010, Steele, et al. 2013) and may therefore be less concerned about neighbourhood factors. On the other hand, rented accommodation, especially social housing, tends to be located in the most deprived areas (Clark, et al. 2013) which limits the chance of a move to an affluent area without a change in tenure. Low mobility within the social housing sector, particularly in high-demand areas in the South, is well documented (e.g. Kearns and Parkes 2003) and previous research has found that social renters have limited opportunities to 'move up' to less-advantaged areas or to maintain residence in better-off areas (Clark, et al. 2013).

Turning to individuals who changed tenure, private renters who made the transition into home-ownership in the last year were no more or less averse to area deprivation than those who remained renting (Figures 1 and 2). However, the push effect of deprivation for new homeowners is in the opposite direction to that for more established owners: for renters who became owners in the last year, higher deprivation in the current area was associated with a reduced probability of leaving that area (Figure 2). This finding may reflect a tendency for renters in deprived areas to stay in the same area, or move to another deprived area, when seeking to buy their own home due to a lack of affordable housing in more prosperous neighbourhoods. Lower housing costs in deprived areas are also likely to explain their attraction for individuals leaving the family home for rented accommodation (Figure 2). Finally, proximity to the current residence is of less importance to households whose move coincides with a tenure transition than for those whose tenure stays the same, regardless of the nature of the change (Table 1).

Unobserved heterogeneity in push and pull effects

Table 2 shows posterior estimates of the variances and correlations between the four individual-level random effects representing time-invariant propensities to stay in the same area (inertia), and sensitivities to deprivation as a push or pull factor in residential location choice and distance from current area as a pull factor. The results presented are from the full model that allows for differential effects of deprivation and distance by the observed household characteristics of Table 1. Random effect variances from the reduced model without household characteristics are also shown. (The correlations for this model have been suppressed.) We find that the addition of covariates leads to a substantial reduction in the between-individual variance in inertia and in the push and pull effects of deprivation. The small variance in the effect of distance from the current area appears unexplained by the household covariates.

After accounting for differential effects of deprivation and distance by observed household characteristics, only the correlation between inertia and the pull effect of distance remains important. This strong, positive correlation suggests that more mobile households have a stronger-than-average preference to remain close to their current neighbourhood when moving house. Put another way, households that move less frequently are prepared to move greater distances when they do so.

4. DISCUSSION

This article has presented a general mixed logit model which makes use of longitudinal data to distinguish, and estimate simultaneously, the push and pull effects of multiple area attributes on residential location choice. An efficient MCMC algorithm is proposed which, together with sampling of the choice set, allows consideration of a larger set of potential destination areas, larger sample size and longer observations period than has been possible in previous research.

Our analysis of household heterogeneity in the effects of neighbourhood deprivation on out-mobility and movers' selection of the destination area suggests that the residential choices of less-advantaged households are severely constrained. We find that low income is associated with a lower probability of moving to a more advantaged neighbourhood while private and social renters are less likely than owner-occupiers to move out of deprived areas. As argued by other authors (e.g. Clark, et al. 2013), such constraints in the housing market lead to increasingly selective migration with disadvantaged households unable to 'move up' to better-off areas, a situation which can only worsen with rising house prices. We also find that even for local moves within labour market areas, the influence of distance of a potential destination from the current residence differs markedly according to household characteristics.

The focus of the analysis presented here is household heterogeneity in the importance placed on two area characteristics, deprivation and distance, in location choices. Another avenue for research would be to compare the push and pull effects of multiple area attributes – such as crime, house prices and school quality – for different types of household. It is also straightforward to extend the model to include random coefficients on interactions between household and area characteristics, $\boldsymbol{x}_{it-1} * \boldsymbol{z}_{rt-1}$ in Equation (4), thus allowing for heterogeneity in the importance placed on an area characteristic $\boldsymbol{z}$ within groups defined by $\boldsymbol{x}$. A consequence of this more general specification is that the covariance between individual choice propensities given by Equations (9) and (10) would depend not only on $\boldsymbol{z}$, but also on (possibly time-varying) household characteristics.

Acknowledgements

This research was supported by two grants from the UK Economic and Social Research Council (RES-062-23-2265 and RES-576-25-0032).

References

Albert, J. H., and Chibb, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669-679.

Allenby, G. (1997), "An Introduction to Hierarchical Bayesian Modeling," *Tutorial Notes, Advanced Research Techniques Forum, American Marketing Association*.

Ben-Akiva, M., and Bowman, J. L. (1998), "Intergration of an Activity-Based Model System and a Residential Location Model," *Urban Studies*, 35, 1131-1153.

Ben-Akiva, M., and Lerman, S. R. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge: MIT Press.

Bergström, L., and van Ham, M. (2010), "Understanding Neighbourhood Effects: Selection Bias and Residential Mobility," *Discussion paper no. 5193*.

Bhat, C. R., Govindarajan, A., and Pulugurta, V. (1998), "Disaggregate Attraction-End Choice Modeling: Formulation and Empirical Analysis," *Transportation Research Record*, 1645, 60-68.

Bhat, C. R., and Guo, J. (2004), "A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modelling," *Transportation Research Part B*, 38, 147-168.

Bolt, G., van Kempen, R., and van Ham, M. (2008), "Minority Ethnic Groups in the Dutch Housing Market: Spatial Segregation, Relocation Dynamics and Housing Policy," *Urban Studies*, 45, 1359-1384.

Browne, W. J., and Draper, D. (2000), "Implementation Issues in the Bayesian Fitting of Multilevel Models," *Computational Statistics*, 15, 391-420.

Browne, W. J., Steele, F., Golalizadeh, M., and Green, M. (2009), "The Use of Simple Reparameterisations in Mcmc Estimation of Multilevel Models with Applications to Discrete-Time Survival Models," *Journal of the Royal Statistical Society A (Statistics in Society)*, 172, 579-598.

Brownstone, D., Bunch, D., and Train, K. (2000), "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles," *Transportation Research Part B*, 34, 315-338.

Bruch, E. E., and Mare, R. D. (2012), "Methodological Issues in the Analysis of Residential Preferences, Residential Mobility, and Neighborhood Change," *Sociological Methodology*, 42, 103-154.

Charlton, C. M. J., et al. (2013), "Stat-JR Version 1.0," *Centre for Multilevel Modelling, University of Bristol & Electronics and Computer Science, University of Southampton*.

Clark, W. A. V. (1992), "Residential Preferences and Residential Choices in a Multiethnic Context," *Demography*, 29, 451-466.

Clark, W. A. V., and Huang, Y. (2003), "The Life Course and Residential Mobility in British Housing Markets," *Environment and Planning A*, 35, 323-339.

Clark, W. A. V., and Ledwith, V. (2007), "How Much Does Income Matter in Neighbourhood Choice?," *Population Research and Policy Review*, 26, 145-161.

Clark, W. A. V., van Ham, M., and Coulter, R. (2013), "Spatial Mobility and Social Outcomes," *Journal of Housing and the Built Environment*, Online first.

Crowder, K. D., and South, S. J. (2008), "Spatial Dynamics of White Flight: The Effects of Local and Extralocal Racial Conditional on Neighborhood out-Migration," *American Sociological Review*, 73, 792-812.

Crowder, K. D., South, S. J., and Chavez, E. (2006), "Wealth, Race and Inter-Neighborhood Migration," *American Sociological Review*, 7, 72-94.

Dibben, C., et al. (2007), "Investigating the Impact of Changing the Weights That Underpin the Index of Multiple Deprivation 2004," Technical.

Domanski, A., and von Haefen, R. H. (2010), "Estimation and Welfare Analysis from Mixed Logit Recreation Demand Models with Large Choice Sets."

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995), "Efficient Parameterizations for Normal Linear Mixed Models," *Biometrika*, 82, 479-488.

Greene, W. H., and Hensher, D. A. (2003), "A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit," *Transportation Research Part B: Methodological*, 37, 681-698.

Hedman, L., van Ham, M., and Manley, D. (2011), "Neighbourhood Choice and Neighbourhood Reproduction," *Environment and Planning A*, 43, 1381-1399.

Hensher, D. A., and Greene, W. H. (2003), "The Mixed Logit Model: The State of Practice," *Transportation*, 30, 133-176.

Hensher, D. A., Rose, J. M., and Greene, W. H. (2005), *Applied Choice Analysis: A Primer*, New York: Cambridge University Press.

Ioannides, Y. M., and Zabel, J. E. (2008), "Interactions, Neighbourhood Selection and Housing Demand," *Journal of Urban Economics*, 63, 229-252.

ISER (2009), *British Household Panel Survey: Waves 1-17, 1991-2008* (6th ed.), University of Essex, Institute for Social and Economic Research [original data producer(s)], Colchester, Essex: UK Data Archive [distributor]. SN: 5151:

Jain, D. C., Vilcassim, N. J., and Chintagunta, P. K. (1994), "A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data," *Journal of Business & Economic Statistics*, 12, 317-328.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. (1998), "Markov Chain Monte Carlo in Practice: A Roundtable Discussion," *American Statistician*, 52, 93-100.

Keane, M., and Wasi, N. (2012a), "Comparing Alternative Models of Heterogeneity in Consumer Choice Behaviour," *Journal of Applied Econometrics*, Online advance publication.

Keane, M. P., and Wasi, N. (2012b), "Estimation of Discrete Choice Models with Many Alternatives Using Random Subsets of the Full Choice Set: With an Application to Demand for Frozen Pizza."

Kearns, A., and Parkes, A. (2003), "Living in and Leaving Poor Neighbourhood Conditions in England," *Housing Studies*, 18, 827-851.

Lee, B. A., Oropesa, R. S., and Kanan, J. W. (1994), "Neighbourhood Context and Residential-Mobility," *Demography*, 31, 249-270.

Lee, B. H. Y., and Waddell, P. (2010), "Residential Mobility and Location Choice: A Nested Logit Model with Sampling of Alternatives," *Transportation*, 37, 587-601.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. J. (2000), "Winbugs - a Bayesian Modelling Framework: Concepts, Structure, and Extensibility," *Statistics and Computing*, 10, 325-337.

Maddala, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics* Cambridge: Cambridge University Press.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. P. Zarembka, New York: Academic Press, pp. 105-142.

McFadden, D. (1978), "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, eds. A. Karlqvist, F. Lundqvist, F. Snickars and J. Weibull, Amsterdam: North-Holland, pp. 75-96.

Nechyba, T. J., and Strauss, R. P. (1998), "Community Choice and Local Public Services: A Discrete Choice Approach," *Regional Science and Urban Economics*, 28, 51-73.

Nerella, S., and Bhat, C. R. (2004), "Numerical Analysis of Effect of Sampling of Alternatives in Discrete Choice Models," *Transportation Research Record*, 1894, 11-19.

Quillian, L. (2003), "How Long Are Exposures to Poor Neighbourhoods? The Long-Term Dynamics of Entry and Exit from Poor Neighbourhoods," *Population Research and Policy Review*, 2, 221-249.

Rabe, B., and Taylor, M. P. (2010), "Residential Mobility, Quality of Neighbourhood and Life Course Events," *Journal of the Royal Statistical Society A (Statistics in Society)*, 173, 531-555.

Revelt, D., and Train, K. (1998), "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," *Review of Economics and Statistics*, 80, 647-657.

South, S. J., and Crowder, K. D. (1997), "Escaping Distressed Neighbourhoods: Individual, Community and Metropolitan Influences," *American Journal of Sociology*, 4, 1040-1084.

South, S. J., and Crowder, K. D. (1998), "Leaving the 'Hood: Residential Mobility between Black, White, and Integrated Neighborhoods," *American Sociological Review*, 63, 17-26.

Steele, F., Clarke, P., and Washbrook, E. (2013), "Modeling Household Decisions Using Longitudinal Data from Household Panel Surveys, with Applications to Residential Mobility," *Sociological Methodology*, 43, 225-276.

Train, K. (2001), "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit," *Working paper, Department of Economics, University of California, Berkeley*.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press.

van Ham, M., and Clark, W. A. V. (2009), "Neighbourhood Mobility in Context: Household Moves and Changing Neighbourhoods in the Netherlands," *Environment and Planning A*, 41, 1442-1459.

Table 1. Estimated coefficients from mixed logit model with household characteristics: Push and pull effects* of deprivation (IMD) and pull effects of distance from the current area of residence. Estimates are posterior means and 95% credible intervals (2.5 and 97.5 percentiles).

	Push effect of IMD ($-\beta$)		Pull effect of IMD (γ_1)		Pull effect of distance (γ_2)	
Household characteristics	Median	95% CI	Median	95% CI	Median	95% CI
Reference group[†]	0.301	(0.121, 0.482)	-0.699	(-0.839,-0.566)	-0.642	(-0.681, -0.604)
log(income)	0.016	(-0.025, 0.058)	-0.044	(-0.068, -0.020)	0.003	(-0.002, 0.008)
Household type $t-1$ (ref = couple no child)						
Single female	-0.101	(-0.353, 0.151)	-0.103	(-0.265, 0.059)	0.019	(-0.013, 0.051)
Single male	-0.111	(-0.342, 0.122)	0.016	(-0.137, 0.169)	0.018	(-0.010, 0.046)
Single parent	-0.475	(-0.753, -0.199)	0.167	(-0.022, 0.355)	-0.038	(-0.076, -0.001)
Couple, y 0-4	-0.155	(-0.390, 0.078)	0.070	(-0.100, 0.238)	0.020	(-0.007, 0.048)
Couple, y 5-10	-0.147	(-0.441, 0.148)	0.045	(-0.182, 0.267)	-0.049	(-0.087, -0.011)
Couple, y 11-15	0.024	(-0.324, 0.371)	0.030	(-0.254, 0.307)	-0.093	(-0.139, -0.048)
Couple, y 16+	0.111	(-0.212, 0.435)	0.303	(0.052, 0.547)	-0.113	(-0.161, -0.067)
Birth $(t-1, t]$	0.342	(0.068, 0.614)	0.321	(0.116, 0.523)	0.026	(-0.007, 0.057)
Housing tenure $t-1$						
(ref = owner)	0	-	0	-	0	-
Private rent	-0.731	(-0.971, -0.497)	0.497	(0.350, 0.644)	0.186	(0.155, 0.218)
Social rent	-0.767	(-1.002, -0.538)	0.826	(0.660, 0.992)	0.024	(-0.018, 0.065)
Living with family	-0.551	(-0.804, -0.301)	0.552	(0.379, 0.730)	-0.009	(-0.046, 0.027)
Tenure change $(t-1, t]$						
Rent → own	-0.192	(-0.447, 0.062)	-0.005	(-0.158, 0.146)	0.363	(0.335, 0.391)
Family → rent	0.065	(-0.275, 0.406)	0.722	(0.528, 0.916)	0.495	(0.453, 0.537)
Own to rent	-0.357	(-0.709, -0.006)	0.551	(0.345, 0.755)	0.542	(0.505, 0.580)
Other change	0.707	(0.417, 1.008)	0.520	(0.339, 0.699)	0.290	(0.254, 0.326)

*The posterior mean of the interia parameter α is 5.948 with a 95% credible interval of (5.815, 6.079).

[†]Push and pull effects for a couple with no children at $t-1$ or birth in $(t-1, t]$, owner-occupiers at $t-1$ with no change in tenure in $(t-1, t]$ and mean log(household income).

Table 2. Estimated random effect variances (diagonal) and correlations (off-diagonal) from mixed logit model with household characteristics. Estimates are posterior means and 95% credible intervals (2.5 and 97.5 percentiles).

	Inertia	Push - IMD	Pull - IMD	Pull - Distance
Inertia	4.363			
	(3.853, 4.910)			
Push - IMD	-0.080	0.686		
	(-0.260, 0.097)	(0.385, 1.019)		
Pull – IMD	-0.210	-0.190	0.099	
	(-0.453, 0.050)	(-0.705, 0.109)	(0.043, 0.175)	
Pull - Distance	0.729	0.041	-0.065	0.129
	(0.689, 0.767)	(-0.089, 0.178)	(-0.248, 0.122)	(0.116, 0.142)
Variances from model without household characteristics				
	5.835	1.084	0.254	0.125
	(5.217, 6.485)	(0.769, 1.423)	(0.170, 0.354)	(0.112, 0.140)

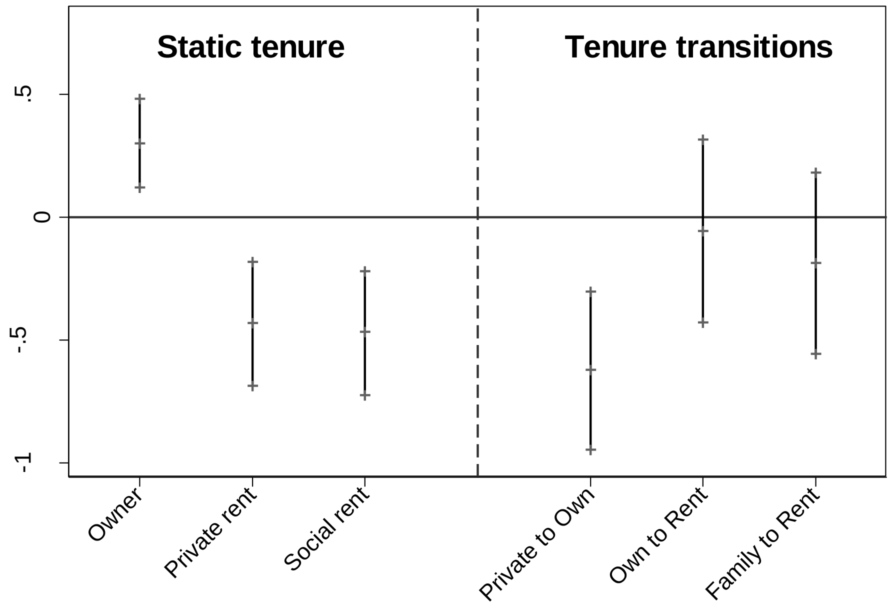

Figure 1. Push effects of deprivation by housing tenure at $t-1$ and t. Estimates are posterior means and 95% credible intervals (2.5 and 97.5 percentiles).

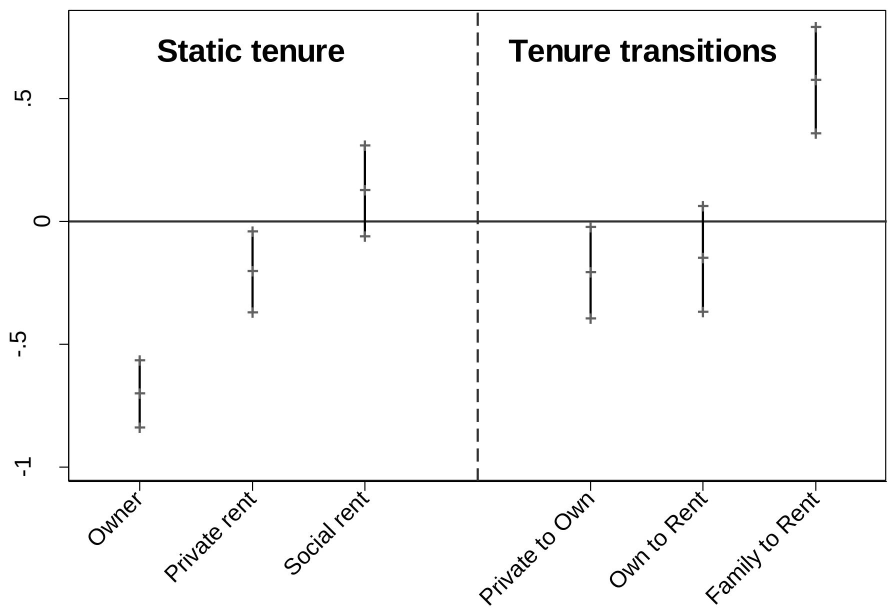

Figure 2. Pull effects of deprivation by housing tenure at $t-1$ and t. Estimates are posterior means and 95% credible intervals (2.5 and 97.5 percentiles).

Appendix: MCMC algorithm for the mixed logit model

Section 2.8 outlines MCMC estimation of a mixed logit with random household-specific coefficients for a subset of predictors. Details of the MCMC algorithm for a more general model with random choice (area)-specific effects v_r are given below.

Consider a model for y_{it}, the observed area of residence at time t for household i where

$$\Pr(y_{it} = r) = p_{rit} = \frac{\exp(\eta_{rit})}{\sum_{k\in\mathcal{C}_{it}} \exp(\eta_{kit})}, \qquad r = 1, \dots R_{it}$$

As described in Section 2.8, the linear predictor for a model with household random effects for a subset of the fixed parameters can be written

$$\eta_{rit} = \boldsymbol{\theta}_i{}^{\boldsymbol{T}} \boldsymbol{A}_{rit} + \boldsymbol{\varphi}^T \boldsymbol{B}_{rit} + v_r - \ln(q_{rit})$$

where

$$\boldsymbol{\theta}_i \sim MVN(\boldsymbol{\theta}, \boldsymbol{\Omega}_u), \qquad \boldsymbol{u}_i = \boldsymbol{\theta}_i - \boldsymbol{\theta}, \qquad \boldsymbol{u}_i \sim MVN(0, \boldsymbol{\Omega}_u)$$

and

$$v_r \sim N(0, \sigma_v^2)$$

Bayesian model estimation requires the addition of prior distributions to parameters not already expressed and so the full posterior distribution can be written:

$$p(\boldsymbol{\theta}_i, \boldsymbol{\theta}, \boldsymbol{\varphi}, v_r, \boldsymbol{\Omega}_u, \sigma_v^2 | y_{it}) \propto \prod_{rit} L(y_{it} | \boldsymbol{\theta}_i, \boldsymbol{\varphi}, v_r) \prod_i p(\boldsymbol{\theta}_i | \boldsymbol{\theta}, \boldsymbol{\Omega}_u) \prod_r p(v_r | \sigma_v^2)\, p(\boldsymbol{\theta})\, p(\boldsymbol{\varphi})\, p(\boldsymbol{\Omega}_u)\, p(\sigma_v^2)$$

We assume diffuse Uniform priors for the fixed effects, denoted by $p(\boldsymbol{\theta})$ and $(\boldsymbol{\varphi})$, and an inverse Gamma(ε,ε) prior for the between-area variance, $p(\sigma_v^2)$. For the between-household variance matrix, with prior $p(\boldsymbol{\Omega}_u)$, we specify a Wishart prior for the precision matrix with parameters v_p and $\boldsymbol{S}_p$. The Uniform prior used in the application is a special case with $v_p = -n_u - 1$ and $\boldsymbol{S}_p = \boldsymbol{0}$, where n_u is the number of predictors with random coefficients at the household level. Browne and Draper (2000) compare prior distributions for variance matrices and show that choice of prior becomes less important for large numbers of random effects, such as in the application where there are over 6000 households.

The MCMC algorithm proceeds by generating draws in turn from the full conditional posterior distributions as follows.

Step 1: The household effects, $\boldsymbol{\theta}_i$, have conditional posterior distributions

$$p(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \boldsymbol{\varphi}, v_r, \boldsymbol{\Omega}_u, y_{it}) \propto \prod_{rt} L(\, y_{it}|\boldsymbol{\theta}_i, \boldsymbol{\varphi}, v_r) p(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \boldsymbol{\Omega}_u)$$

As this does not have a standard form, we use a random walk Metropolis algorithm with univariate Normal proposal distributions and an adaptive method for tuning the proposal variance as described in Browne, et al. (2000).

Step 2: The fixed effects, $\boldsymbol{\theta}$, have conditional posterior distribution

$$p(\boldsymbol{\theta}|\boldsymbol{\Omega}_u) \sim N_{n_u}\left(\sum\nolimits_{i=1}^{n} \frac{\boldsymbol{\theta}_i}{n}, \frac{\boldsymbol{\Omega}_u}{n}\right)$$

where n is the total number of households. We can sample directly from this distribution using Gibbs sampling.

Step 3: The household-level variance matrix, $\boldsymbol{\Omega}_u$, has associated precision matrix $\boldsymbol{\Omega}_u^{-1}$ which has conditional posterior distribution

$$p(\Omega_u^{-1}|\boldsymbol{\theta}_i, \boldsymbol{\theta}) \sim W(n + v_p, \sum\nolimits_{i=1}^{n} (\boldsymbol{\theta}_i - \boldsymbol{\theta})^T (\boldsymbol{\theta}_i - \boldsymbol{\theta}) + \boldsymbol{S}_p)$$

This step again uses Gibbs sampling.

Step 4: The other fixed effects, $\boldsymbol{\varphi}$, have conditional posterior distributions

$$p(\boldsymbol{\varphi}|\boldsymbol{\theta}_i, v_r, y_{it}) \propto \prod_{rit} L(\, y_{it}|\boldsymbol{\theta}_i, \boldsymbol{\varphi}, v_r)\, p(\boldsymbol{\varphi})$$

As in Step 1, a random walk Metropolis algorithm with univariate Normal proposal distributions and an adaptive method for tuning the proposal variance is used.

Step 5: The area-level random effects, v_r, have conditional posterior distributions

$$p(v_r|\boldsymbol{\theta}_i, \boldsymbol{\varphi}, y_{it}, \sigma_v^2) \propto \prod_{it} L(\, y_{it}|\boldsymbol{\theta}_i, \boldsymbol{\varphi}, v_r)\, p(v_r|\sigma_v^2)$$

This step uses the same sampling method as Steps 1 and 4.

Step 6: The variance of the area effects, σ_v^2

$$p(\sigma_v^2|v_r) \sim \Gamma^{-1}(\varepsilon + \frac{R}{2}, \varepsilon + \frac{1}{2}\sum_{r=1}^{R} v_r^2)$$

where R is the total number of possible choices (areas of residence) across all travel-to-work areas, individuals and years. This step again uses Gibbs sampling.

Note that to fit a non-hierarchically centred formulation of the model, the fixed effects, $\boldsymbol{\theta}$, are considered as part of $\boldsymbol{\varphi}$ by incorporating the data matrix $\boldsymbol{A}_{rit}$ into $\boldsymbol{B}_{rit}$ and then replacing $\boldsymbol{\theta}$ with $\mathbf{0}$ in the prior for $\boldsymbol{\theta}_i$. To perform orthogonal parameterisation (as detailed in Browne, et al. (2009)) the matrix $\boldsymbol{B}_{rit}$ is converted into a matrix of orthogonal vectors and the algorithm is run using this matrix. The chains for the parameters $\boldsymbol{\varphi}$ are then post-processed to obtain parameters for the original parameterisation.