Neighbourhood choice of young families in England

Elizabeth Washbrook and Fiona Steele

Recent Advances in Research on Housing Transitions and the Life Course Symposium

10th December 2013, Royal Statistical Society

What do households value when choosing a neighbourhood?

Challenging research question

- Neighbourhoods are multi-dimensional
- Area characteristics must be chosen together as a "package" from a limited set of alternatives
- Classic problem in study of US racial segregation

A **longitudinal discrete-choice approach** allows us to disentangle which area characteristics

- Attract inward migration (**pull** factors)
- Promote outward migration (**push** factors)

Understanding neighbourhood choice

- Crucial determinant of the dynamics of **neighbourhood reproduction and change**, and of patterns of spatial segregation
- Creates selection bias in models of the causal effect of place on people
  - Models of the processes behind the sorting of people into neighbourhoods should be "an integral part of the conceptual framework we use to understand **neighbourhood effects**" Bergström and van Ham (2010)
- Aspect of equality of opportunity in terms of **upward socio-spatial mobility** over the life course

The “choice” problem

- Each year a household chooses a neighbourhood from their current labour market area (**the choice set**) in which to live
  - The chosen neighbourhood may be the same one as the previous year (no move) or an alternative
- Aim to isolate the degree a particular area characteristic attracts or repels young families when other characteristics are held constant
- Need to be critical of the concept of choice in this setting
  - Location choices reflect constraints as well as preferences
  - Interpret results as description of a **multi-dimensional matching process** (Quillian, 2013)

Application to young families

- Study of locational choices of female household members from shortly before the birth of the first child to the child's 12th birthday
- Life course literature predicts certain area characteristics will matter disproportionately during this period (Mulder, 2006)
- Large literature on neighbourhood effects on children (eg Leventhal and Brooks-Gunn, 2000) raises issues of intergenerational equity
- Focus on whether push and pull factors differ with socio-economic status (maternal education)

Area definitions

TTWA map

The choice set is restricted to the current **Travel-to-Work-Area (TTWA)**

166 TTWAs (at least partly) in England

A neighbourhood is a **lower super output area (LSOA)**

32,482 LSOAs in England, average 1500 residents

Number of LSOAs per TTWA ranges from 4 to 5647 (London), mean =196

12.5% sample annual cross-LSOA mobility rate. **77%** of all moves within-TTWA, another **13%** just across boundary.

LSOA characteristics

A. Dwelling characteristics

HOMESIZE: Average number of rooms (excl bathrooms) per occupied household space. (2001 Census). Mean = 5.2, range = (2.4, 7.6)

HOMEQUAL: Average value of dwelling in 1991 (from council tax bands) divided by HOMESIZE. Mean = 12.6, range = (3.5, 75.3) (Thousands 1991GBP; logged in analysis)

OWNRATE: Proportion of dwellings owner-occupied. (2001 Census). Mean = 0.69, range = (0.02, 1)

LSOA characteristics

B. Physical environment

AREA: Size of LSOA in thousands of sq km. (Captures population density.) Mean = 4.1, range = (0.02, 685)

GREEN: Proportion of land area classed as green space. Mean = 0.43, range = (0, 0.99)

OUTDOOR: IMD Outdoor Living Environment sub-domain z-score. Combines indicators of air quality and road traffic accidents. Mean (2007) = 0, range = (-1.1, 3.9)

AMENITIES: IMD Barriers to Access sub-domain z-score. Combines indicators of road distance to GP, grocery, primary school and post office. Mean (2007) = 0, range = (-3.7, 4.6)

LSOA characteristics

C. SOCIAL ENVIRONMENT

EDUCATION: IMD Children and Young People's Skills sub-domain z-score. Mean (2007) = 0, range = (-1.1, 3.9)

CRIME: IMD Crime domain z-score. Mean (2007) = 0, range = (-4.5, 3.8)

CHILDRATE: Proportion of households with dependent children. (2001 Census). Mean = 0.30, range = (0.01, 0.85)

D. **DISTANCE**: Distance from current LSOA of residence in km. (Uses population-weighted LSOA centroids).

LSOA-level correlations

	HOME SIZE	HOME QUAL	OWN RATE	AREA	GREEN	OUT DOOR	AMENITIES	EDUCATION	CRIME	CHILD RATE
HOMESIZE	1									
HOMEQUAL	0.29	1								
OWNRATE	**0.74**	0.35	1							
AREA	0.29	0.15	0.08	1						
GREEN	0.43	0.14	0.25	0.46	1					
OUTDOOR	**0.56**	0.02	0.49	0.24	0.50	1				
AMENITIES	**-0.54**	-0.24	-0.37	-0.50	**-0.63**	**-0.54**	1			
EDUCATION	**0.51**	**0.68**	**0.66**	0.16	0.21	0.26	-0.29	1		
CRIME	**0.53**	0.41	**0.53**	0.30	0.39	**0.58**	-0.50	**0.56**	1	
CHILDRATE	0.24	-0.22	0.05	-0.07	-0.00	0.08	-0.06	-0.20	-0.01	1
INCOMES	0.62	0.56	0.79	0.17	0.30	0.47	-0.40	0.76	0.60	-0.23

Signs of correlations chosen so that higher scores always reflect area improvements.
Time-varying characteristics shown at 2004 values. N=32,482 LSOAs

Modeling approach

U_{ir} is the unobserved **utility**, or psychic benefit, household i would get from living in LSOA r.

Y_{it} is the LSOA **actually chosen** by the household at time t. We assume the household chooses the LSOA that yields the highest possible U_{ir}.

U_{ir} depends on

- The **characteristics** of the LSOA, $Z_{r1}, \ldots, Z_{rJ}$
- Individual responses to those characteristics
- Whether the household is currently resident in area r at $t-1$ (Y_{it-1})

$$U_{ir} = \gamma_{i1} Z_{r1} + \gamma_{i2} Z_{r2} + \cdots \gamma_{iJ} Z_{rJ} \qquad \text{if } Y_{it-1} \neq r \quad \text{(new LSOA)}$$

$$U_{ir} = \alpha_i + \beta_{i1} Z_{r1} + \beta_{i2} Z_{r2} + \cdots \beta_{iJ} Z_{rJ} \qquad \text{if } Y_{it-1} = r \quad \text{(home LSOA)}$$

From utilities to probabilities

$$U_{ir} = \gamma_{i1} Z_{r1} + \gamma_{i2} Z_{r2} + \cdots \gamma_{iJ} Z_{rJ} \qquad \text{if } Y_{it-1} \neq r$$

$$U_{ir} = \alpha_i + \beta_{i1} Z_{r1} + \beta_{i2} Z_{r2} + \cdots \beta_{iJ} Z_{rJ} \qquad \text{if } Y_{it-1} = r$$

Multinomial logit model for probability LSOA r is selected:

$$\Pr(Y_{it} = r) = \frac{\exp(U_{ir})}{\sum_k \exp(U_{ik})}$$

The probability of selecting r relative to the probability of selecting s:

$$\frac{\Pr(Y_{it} = r)}{\Pr(Y_{it} = s)} = \exp(U_{ir} - U_{is})$$

E.g. for $Y_{it-1} \neq r$ or s:

$$\frac{\Pr(Y_{it} = r)}{\Pr(Y_{it} = s)} = \exp\{\gamma_{i1}(Z_{r1} - Z_{s1}) + \cdots + \gamma_{iJ}(Z_{rJ} - Z_{sJ})\}$$

The **pull** effect of Z_j

$\exp(\gamma_{ij})$ is the effect of a 1 unit ↑ in Z_j on the relative probability of **moving into** an LSOA, holding all else constant.

If $\exp(\gamma_{ij}) = 1$, Z_j has no effect on the probability of a move into a new area (in-migration).

If $\exp(\gamma_{ij}) < 1$, higher Z_j reduces the risk of move to an area.

If $\exp(\gamma_{ij}) > 1$, higher Z_j increases the risk of move to an area.

The **push** effect of Z_j

$1/\exp(\beta_{ij})$ is the effect of a 1 unit ↑ in Z_j on the relative probability of **moving out of** one's current LSOA, holding all else constant.

If $1/\exp(\beta_{ij}) = 1$, Z_j has no effect on the probability of a moving away from current LSOA (out-migration).

If $1/\exp(\beta_{ij}) < 1$, higher Z_j reduces the risk of leaving.

If $1/\exp(\beta_{ij}) > 1$, higher Z_j increases the risk of leaving.

Individual heterogeneity

Pull (and push) effect of each Z_j depends on **observable individual characteristics** (X_i) like socio-economic status...

$$\gamma_{ij} = \gamma_{0ij} + \gamma_{1j} X_i$$

...and on an **individual random effect** that is uncorrelated with X_i and that is normally distributed among the population.

$$\gamma_{0ij} \sim N(\gamma_{0j}, \sigma^2_{\gamma j})$$

Random effects or "tastes" for different area characteristics can be correlated within households. We might expect a positive correlation in the push and pull effects of a given characteristic.

The inclusion of random effects in the model relaxes the **Independence of Irrelevant Alternatives** (IIA) property of the multinomial logit model.

Individual-level data

- Longitudinal data on women from the BHPS, waves 9-18 (1999 to 2008), living in known LSOA in England at *t*-1 and *t*
- Inclusion criteria for main sample: **age of oldest child** at *t*-1 is between -2 and 11
- Explore differences in decision process by **maternal education**:
  - **High** = Degree/teaching qual (**23%**); **Medium** = A-level/other higher qual (eg nursing) (**39%**); **Low** = GCSE/O-level or below (**39%**)
- Records with a cross-TTWA move between *t*-1 and *t* dropped, *except* for moves within 43km. (867 moves; 11%)
- 1543 individuals, 7589 person-years; 6.5 million person-wave-LSOA records!

Handling large choice sets

A dataset of 6.5 million observations is impractical

We randomly sample neighbourhoods from within an individual's choice set with probability q_{rit}

- q_{rit} = 1 for areas of residence at t and t -1, else q_{rit} << 1
- Include –log(q_{rit}) as offset term in model
- No effect of sampling on consistency of parameter estimates from model (McFadden 1978)

Sample LSOAs with probability proportional to the square root of number of LSOAs in TTWA → 389,018 person-wave-LSOA records

Estimation

- Implemented using MCMC methods in computationally-efficient software (Stat-JR)
  - With big thanks to Bill Browne and Chris Charlton!
  - Handles the 18 random effects (170 RE parameters)
  - 5 chains with different random starts run in parallel
  - 50k iterations (with 5000 burn-in), estimation time currently 9 hours
- Evidence of substantial individual unobserved heterogeneity in responses to area characteristics
  - DIC for single-level model = 14651
  - DIC for random effects model = 8627

Results: Pull effects for households with REs at mean (relative probability of in-mobility)

Z (CHANGE)	High ed	Medium ed	Low ed	Equality constraints
HOMESIZE (+1)	**3.46** (1.48,8.62)	1.22 (0.72,2.13)		M = L
HOMEQUAL (+10%)	**0.85** (0.74,0.96)	**0.92** (0.86,0.99)		M = L
OWNRATE (+10ppt)	1.13 (0.96,1.31)			H = M = L
AREA (+1000 SQ KM)	**0.85** (0.78,0.91)	**0.85** (0.78,0.91)		H = M = L
GREEN (+10ppt)	**1.26** (1.08,1.54)	1.10 (0.99,1.22)		M = L
OUTDOOR (-1 SD)		**1.70** (1.21,2.33)		H = M = L
AMENITIES (-1 SD)	1.53 (0.92,2.73)	0.85 (0.64,1.17)		M = L
EDUCATION (-1 SD)	1.58 (0.90,2.95)	0.91 (0.63,1.34)	0.55 (0.37,0.81)	
CRIME (-1 SD)	1.32 (0.80,2.27)	**1.50** (1.02,2.26)	1.16 (0.81,1.84)	
CHILDRATE (+ 5ppt)	**0.73** (0.57,0.89)	1.04 (0.94,1.14)		M = L
DISTANCE (- 1KM)	**2.33** (1.96,3.20)	**2.77** (2.40,3.95)		M = L

Numbers are the probability ratio associated with an **improvement** in the LSOA characteristic of the specified magnitude (summaries over 50k chains). 95% credible intervals in parentheses. Equality constraints after pre-testing on single level model. N = 389,018 person-wave-LSOAs.

Results: Individual heterogeneity in pull effects (estimates at +/- 1 SD of random effect)

Z (CHANGE)	High ed	Medium ed	Low ed
HOMESIZE (+1)	(0.28, 43.3)	(0.10, 15.3)	
HOMEQUAL (+10%)	na		
OWNRATE (+10ppt)	(0.51, 2.52)		
AREA (+1000 SQ KM)	(0.70, 1.04)		
GREEN (+10ppt)	(0.71, 2.26)	(0.61, 1.96)	
OUTDOOR (-1 SD)	(0.28, 10.2)		
AMENITIES (-1 SD)	(0.29, 8.06)	(0.16, 4.49)	
EDUCATION (-1 SD)	(0.34, 7.34)	(0.20, 4.22)	(0.12, 2.56)
CRIME (-1 SD)	(0.24, 7.19)	(0.28, 8.17)	(0.21, 6.31)
CHILDRATE (+ 5ppt)	na		
DISTANCE (- 1KM)	(1.10, 4.95)	(1.31, 5.88)	

Numbers give an indication of the range in probability ratios across individuals associated with the specified improvement in Z_j. na = random effect dropped from model on the basis of the DIC.

Results: Push effects for households with median RE (relative probability of out-mobility)

Z (CHANGE)	High ed	Medium ed	Low ed	Equality constraints
HOMESIZE (+1)	0.48 (0.13,1.34)			H = M = L
HOMEQUAL (+10%)	1.14 (0.89,1.50)	0.99 (0.85,1.17)	1.05 (0.91,1.24)	
OWNRATE (+10ppt)	0.83 (0.57,1.17)	0.72 (0.53,1.02)	**0.63** (0.47,0.85)	
AREA (+1000 SQ KM)	**0.75** (0.62,0.90)			H = M = L
GREEN (+10ppt)	0.87 (0.68,1.05)			H = M = L
OUTDOOR (-1 SD)	0.44 (0.12,1.34)	1.00 (0.44,2.59)	1.49 (0.61,4.29)	
AMENITIES (-1 SD)	**0.28** (0.08,0.68)		**0.38** (0.12,0.87)	H = M
EDUCATION (-1 SD)	**0.28** (0.06,0.96)	1.82 (0.85,4.23)		M = L
CRIME (-1 SD)	0.90 (0.39, 1.82)			H = M = L
CHILDRATE (+ 5ppt)	**0.76** (0.48,0.99)			H = M = L

Numbers are the probability ratio associated with an **improvement** in the LSOA characteristic of the specified magnitude (summaries over 50k chains). 95% credible intervals in parentheses. Equality constraints after pre-testing on single level model. N = 389,018 person-wave-LSOAs.

Results: Individual heterogeneity in push effects (estimates at +/- 1 SD of random effect)

Z (CHANGE)	High ed	Medium ed	Low ed
HOMESIZE (+1)	(0.002, 91.4)		
HOMEQUAL (+10%)	na		
OWNRATE (+10ppt)	(0.40,1.72)	(0.35,1.50)	(0.30, 1.32)
AREA (+1000 SQ KM)	(0.41, 1.39)		
GREEN (+10ppt)	(0.27, 2.79)		
OUTDOOR (-1 SD)	(0.005,37.8)	(0.012,86.1)	(0.017,128)
AMENITIES (-1 SD)	(0.002,49.6)		(0.002,65.4)
EDUCATION (-1 SD)	(0.008,9.15)	(0.06, 59.5)	
CRIME (-1 SD)	(0.013, 63.4)		
CHILDRATE (+ 5ppt)	na		

Numbers give an indication of the range in probability ratios across individuals associated with the specified improvement in Z_j. na = random effect dropped from model on the basis of the DIC.

Selected random effect correlations I: Tastes for a given area characteristics in current and alternative locations

Z (LSOA type)	Z (LSOA type)	Corr (95% cred int)
HOMESIZE (home)	HOMESIZE (alternative)	**0.27** (0.02, 0.47)
OWNRATE (home)	OWNRATE (alternative)	0.29 (-0.08, 0.46)
AREA (home)	AREA (alternative)	0.08 (-0.21, 0.31)
GREEN (home)	GREEN (alternative)	**0.48** (0.24, 0.71)
OUTDOOR (home)	OUTDOOR (alternative)	0.22 (-0.06, 0.43)
AMENITIES (home)	AMENITIES (alternative)	**0.52** (0.35, 0.67)
EDUCATION (home)	EDUCATION (alternative)	**0.30** (0.08, 0.47)
CRIME (home)	CRIME (alternative)	**0.43** (0.24, 0.59)

Selected random effect correlations II

Z (LSOA type)	Z (LSOA type)	Corr (95% cred int)
INERTIA (home)	OWNRATE (alternative)	**0.60** (0.43, 0.72)
HOMESIZE (home)	DISTANCE (alternative)	**0.36** (0.15, 0.55)
HOMESIZE (alternative)	DISTANCE (alternative)	**0.34** (0.17, 0.52)
GREEN (home)	OUTDOOR (home)	**0.17** (0.01, 0.43)
HOMESIZE (alternative)	EDUCATION (alternative)	**-0.33** (-0.53, -0.10)
OWNRATE (alternative)	EDUCATION (alternative)	**-0.35** (-0.52, -0.15)

Extensions I: Are young families different from childless women? Selected coefficients

Z (CHANGE) × ED GROUP	Young families	Childless women
Pull effects (in-mobility)		
HOMESIZE (+1) × HIGH	**3.46** (1.48,8.62)	1.19 (0.74,1.90)
HOMEQUAL (+10%) × HIGH	**0.85** (0.74,0.96)	0.97 (0.91,1.03)
GREEN (+10ppt) × HIGH	**1.26** (1.08,1.54)	0.94 (0.81,1.05)
DISTANCE(-1 KM) × HIGH	**2.33** (1.96,3.20)	**1.76** (1.59,1.96)
DISTANCE (-1 KM) × MED/LOW	**2.77** (2.40,3.95)	**1.98** (1.79,2.22)
Push effects (out-mobility)		
INERTIA *[baseline mobility rate]*	**0.006** (.0004,0.4)	**0.011** (.001,0.12)
AMENITIES (-1 SD) × HI/MED	**0.28** (0.08,0.68)	1.64 (0.87, 3.58)
AMENITIES (-1 SD) × LOW	**0.38** (0.12,0.87)	1.44 (0.74, 2.75)
EDUCATION (-1 SD) × HIGH	**0.28** (0.06,0.96)	1.63 (0.60,4.34)
EDUCATION (-1 SD) × M/LOW	1.82 (0.85,4.23)	**2.64** (1.30,5.38)

Extensions II: Are choices different pre- and post-birth? Selected coefficients

Z (CHANGE) × ED GROUP	1 or 2 years before 1st birth	0-11 years after 1st birth	95% CI of diff contains zero?
Pull effects (in-mobility)			
HOMESIZE (+1) × HIGH	4.46 (0.98,34.66)	**4.07** (1.44,14.76)	YES
HOMEQUAL (+10%) × HIGH	**0.61** (0.42,0.80)	0.90 (0.75,1.06)	**NO**
GREEN (+10ppt) × HIGH	1.02 (0.69,1.42)	**1.47** (1.18,2.15)	**NO**
DISTANCE(-1 KM) × HIGH	**3.05** (1.98,7.14)	**2.66** (1.93,5.42)	YES
DISTANCE (-1 KM) × MED/LOW	**3.91** (2.35,6.97)	**3.38** (2.46,7.45)	YES
Push effects (out-mobility)			
INERTIA *[baseline mobility rate]*	0.27 (<0.001, 1038)	**0.0004** (<.0001,0.04)	**NO**
EDUCATION (-1 SD) × HIGH	**0.01** (0.00,0.09)	0.61 (0.11,5.21)	**NO**
EDUCATION (-1 SD) × M/LOW	1.12 (0.18,8.77)	1.97 (0.72,8.23)	YES
CHILDRATE (+ 5ppt)	**0.36** (0.16,0.62)	0.81 (0.58, 1.09)	**NO**

Conclusions

- The matching of young families to neighbourhoods is multi-dimensional: a family must balance desires for multiple characteristics when selecting a location
- Considerable individual variation in responsiveness of choice to different area characteristics, including systematic differences by maternal education
- On average, young families are attracted to neighbourhoods with more green space, better environmental quality and closer proximity to current location, and tend to leave neighbourhoods with poorer access to amenities
- High-SES families
  - make larger trade-offs of dwelling quality/value in exchange for dwelling size
  - are more likely to leave areas with poorer educational outcomes and are more sensitive to education when choosing a new location than low SES families

Next steps

- Substantive
  - Can observed characteristics explain individual heterogeneity?
  - E.g. births of subsequent children, housing tenure
  - Alternative definitions of SES, e.g. income
  - Better measures of school quality?
- Methods
  - Currently investigating methods to improve MCMC efficiency
  - Hierarchical centring for estimation of large RE variances