Steele, F. A., Washbrook, E., Charlton, C., & Browne, W. J. (2016). A Longitudinal Mixed Logit Model for Estimation of Push and Pull Effects in Residential Location Choice. *Journal of the American Statistical Association*, *111*(515), 1061-1074. https://doi.org/10.1080/01621459.2016.1180984

Supplementary materials for "A longitudinal mixed logit model for estimation of push and pull effects in residential location choice"

Fiona Steele,* Elizabeth Washbrook,† Christopher Charlton,† William J. Browne†

1 MCMC algorithm for the mixed logit model

Section 2.8 outlines MCMC estimation of a mixed logit model with random household-specific coefficients for a subset of predictors. Details of the MCMC algorithm for a more general model with area-level random effects are given here. In the following, the area random effects are defined for each choice (neighbourhood), v_r. However, the algorithm is the same when, as in the application of the paper, random effects are specified for higher-level geographical units in which neighbourhoods are nested, $v_{r(d)}$.

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

$$\Pr(y_{it} = r|\boldsymbol{\eta}_{it}) = p_{rit} = \frac{\exp(\eta_{rit})}{\sum_{k \in \mathcal{C}_{it}} \exp(\eta_{kit})}, \quad r = 1, \ldots, R_{it}$$

The linear predictor for a model with household random effects for a subset of the fixed parameters and area effects can be written

$$\eta_{rit} = \boldsymbol{\theta}_i^T \mathbf{A}_{ri(t-1)} + \boldsymbol{\varphi}^T \mathbf{B}_{ri(t-1)} + v_r - \log(q_{rit})$$

where

$$\boldsymbol{\theta}_i \sim \text{MVN}(\boldsymbol{\theta}, \boldsymbol{\Omega}_u), \quad \mathbf{u}_i = \boldsymbol{\theta}_i - \boldsymbol{\theta}, \quad \mathbf{u}_i \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Omega}_u), \quad v_r \sim \text{N}(0, \sigma_v^2)$$

*Department of Statistics, London School of Economics and Political Science
†Centre for Multilevel Modelling, Graduate School of Education, University of Bristol

Bayesian model estimation requires the specification of prior distributions for all parameters 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 $p(\boldsymbol{\varphi})$, 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 ν_p and $\mathbf{S}_p$. The Uniform prior used in the application is a special case with $\nu_p = -n_u - 1$ and $\mathbf{S}_p = 0$, where n_u is the number of predictors with random coefficients at the household level. Specifying a Uniform prior has the advantage of not having to construct a prior guess for $\mathbf{S}_p = 0$ that may need scaling for particular applications. Browne and Draper (2000) compare prior distributions for variance matrices and show that the choice of prior becomes less important for large numbers of random effects, such as in our 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

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

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 and Draper (2000).

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

$$p(\boldsymbol{\theta}|\boldsymbol{\Omega}_u) \sim \mathrm{N}_{n_u}\left(\sum_{i=1}^{n} \frac{\boldsymbol{\theta}_i}{n}, \frac{\boldsymbol{\Omega}_u}{n}\right)$$

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, $\mathbf{\Omega}_u$, has associated precision matrix $\mathbf{\Omega}_u^{-1}$ which has conditional posterior distribution

$$p(\mathbf{\Omega}_u^{-1}|\boldsymbol{\theta}_i, \boldsymbol{\theta}) \sim \mathrm{W}\left(n + \nu_p, \sum_{i=1}^{n}(\boldsymbol{\theta}_i - \boldsymbol{\theta})^T(\boldsymbol{\theta}_i - \boldsymbol{\theta}) + \mathbf{S}_p\right)$$

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 has conditional posterior

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

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 vector $\mathbf{A}_{ri(t-1)}$ into $\mathbf{B}_{ri(t-1)}$ and then replacing $\boldsymbol{\theta}$ with $\mathbf{0}$ in the prior for $\boldsymbol{\theta}_i$. To perform orthogonal parameterisation $\mathbf{B}_{ri(t-1)}$ is transformed to an orthogonal vector and the algorithm is run using this vector (see Browne et al. 2009). The chains for the parameters $\boldsymbol{\varphi}$ are then post-processed to obtain parameters for the original parameterisation.

2 Simulation study

A simulation study was conducted to explore the impact of sample size (number of households n) and panel length T on the bias and accuracy of estimates of the regression coefficients and household random effect parameters. We also investigated the extent to which the correct random effects covariance structure was recovered by fitting models to simulated data with correlated and independent household random effects.

Data generating model

Denote by y_{it} the categorical response indicating the area of residence for household i at wave t. In the data generating model, we make the simplifying assumption that each household has a choice set $\mathcal{C}_{it}$ of size $R_{it} = 45$ at each wave. The decision to move and the choice of location depend on an area characteristic z_r and its interaction with a household characteristic x_i. These are in turn interacted with $w_{ri(t-1)}$, a binary indicator of whether household i is resident in area r at wave $t-1$, and its complement $1 - w_{ri(t-1)}$, to separate push and pull effects. The model additionally allows for a main effect of x_i on the decision to move.

The data generating model takes the form

$$\Pr(y_{it} = r) = \frac{\exp(\eta_{rit})}{\sum_{k \in \mathcal{C}_{it}} \exp(\eta_{kit})}, \quad r = 1, \ldots, 45 \tag{A1}$$

where for $t > 1$ the linear predictor η_{rit} is

$$\begin{aligned} \eta_{rit} &= w_{ri(t-1)}(\alpha_{0i} + \alpha_1 x_i + \beta_{0i} z_r + \beta_1 x_i z_r) \\ &+ (1 - w_{ri(t-1)})(\gamma_{0i} z_r + \gamma_1 x_i z_r) \end{aligned}$$

and

$$\begin{aligned} \alpha_{0i} &= \alpha_0 + u_{\alpha i} \\ \beta_{0i} &= \beta_0 + u_{\beta i} \\ \gamma_{0i} &= \gamma_0 + u_{\gamma i} \end{aligned}$$

where $\mathbf{u}_i = (u_{\alpha i}, u_{\beta i}, u_{\gamma i})$ are normally distributed household random effects with covariance matrix

$$\mathbf{\Omega}_u = \begin{pmatrix} \sigma_\alpha^2 & & \\ \sigma_{\alpha\beta} & \sigma_\beta^2 & \\ \sigma_{\alpha\gamma} & \sigma_{\beta\gamma} & \sigma_\gamma^2 \end{pmatrix}$$

Random effects were generated from a multivariate normal distribution parameterised by the variances $(\sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2)$ and correlations $(\rho_{\alpha\beta}, \rho_{\alpha\gamma}, \rho_{\beta\gamma})$. The random effects allow for unobserved heterogeneity between households in mobility $(u_{\alpha i})$, the push effect of z_r $(u_{\beta i})$ and the pull effect of z_r $(u_{\gamma i})$.

A household's initial area of residence at $t = 1$ was generated from a multinomial logit model with the same form as Equation (A1), but with linear predictor $\eta_{ri1} = v_r$ where $v_r \sim N(0, \sigma_v^2)$ is an area-level random effect.

The parameter values specified for the data generating model are based on parameter estimates from fitting a simple model of the same form as Equation (A1) to the British Household Panel Survey (BHPS) data, with push and pull effects of deprivation (z_r) and its interaction with log-income (x_i). Both variables are time-varying in the application, but treated as time-invariant in the simulations. The parameter values were also chosen to give a population-averaged annual mobility rate of 10%, as in the BHPS data. The parameter values used to generate the data are given in Tables A1 and A2 (under 'True').

Simulation conditions

Five simulation conditions were considered to explore the impact on bias and accuracy of (i) increasing the number of waves T, (ii) increasing the number of households n, and (iii) incorrect specification of the random effects correlation structure. The number of waves simulated was in fact $T + 1$, but the model includes wave 1 information only to define current location and lagged covariates as predictors for residential choice at wave 2. Datasets were simulated for 5, 10 and 15 waves. In the BHPS data, households were observed for up to 10 waves, but the mean number of waves per household was 4.9 due to a combination of late entrants and attrition. As model estimation is computationally intensive, we consider substantially smaller sample sizes of 1000 and 1500 than the 6249 households in the BHPS sample.

The simulation conditions were as follows:

(A) $n = 1000$, $T = 5$, correlated random effects ($\rho \neq 0$)

(B) $n = 1000$, $T = 10$, correlated random effects ($\rho \neq 0$)

(C) $n = 1000$, $T = 15$, correlated random effects ($\rho \neq 0$)

(D) $n = 1500$, $T = 10$, correlated random effects ($\rho \neq 0$)

(E) $n = 1000$, $T = 10$, independent random effects ($\rho = 0$)

Due to the computational time required to fit each model, the study was limited to 100 replications for each condition. For each replicate, model (A1) was fitted and inferences based on five parallel chains of 5,000 MCMC iterations, each with a different starting value and a burn-in sample of 2,000. The number of iterations and burn-in were respectively 100,000 and 10,000 for the BHPS application, but for the simulated data increasing both had little impact on summary statistics computed from the chains.

Results

Table A1 shows the mean of each parameter estimate across the 100 replications, where a parameter estimate for a given replicate is computed as the posterior mean of the five parallel MCMC chains for the parameter. From a comparison of the results for conditions A, B and C we find that, as expected, bias decreases as the number of waves per household increases. The number of waves has most impact on the inertia variance (σ^2_α), the largest of the random effect variances. However, even in the case of 1000 household and 5 waves, the biases are not large (and are in line with posterior mean estimates from a simulation study by Browne and Draper (2000) that assessed MCMC estimation of random effects binary logistic models with a uniform prior). Increasing the number of households by 50% also leads to bias reduction (from a comparison of conditions B and D). Finally, when a model with correlated random effects is fitted to data generated from a model with independent random effects, the correct correlation structure is recovered (conditions B versus E).

To assess the accuracy of estimation, two summary statistics were computed for each parameter: the mean standard error across replicates (where the SE for a given replicate is the standard deviation of the posterior standard deviation of the MCMC chains), and the standard deviation of the parameter estimate across replicates (the empirical SE). Both measures are presented in Table A2. We find substantial improvements in accuracy with increases in either the number of households or the number of waves, with the largest impact observed for the random effect parameters and in particular σ^2_α. For the regression coefficients, the mean SE and empirical SE are very similar for all scenarios considered, and there is minimal loss in accuracy when the random effects correlation structure is incorrectly specified (condition E).

Table A1: Simulation results from 100 replications. Mean parameter estimates across replicates.

Parameter	True	A $n = 1000$ $T = 5$ $\rho \neq 0$	B $n = 1000$ $T = 10$ $\rho \neq 0$	C $n = 1000$ $T = 15$ $\rho \neq 0$	D $n = 1500$ $T = 10$ $\rho \neq 0$	E $n = 1000$ $T = 10$ $\rho = 0$
Coefficients						
α_0	7.145	7.251	7.194	7.176	7.165	7.165
α_1	0.209	0.193	0.221	0.214	0.219	0.219
β_0	0.057	0.063	0.060	0.060	0.069	0.047
β_1	-0.114	-0.121	-0.113	-0.107	-0.109	-0.122
γ_0	0.144	0.130	0.148	0.136	0.137	0.145
γ_1	-0.103	-0.102	-0.098	-0.104	-0.106	-0.100
Random effect parameters						
σ^2_α	4	4.523	4.323	4.196	4.140	4.344
σ^2_β	1	1.342	1.163	1.110	1.093	1.144
σ^2_γ	0.2	0.232	0.219	0.220	0.212	0.223
$\rho_{\alpha\beta}$	-0.15/0	-0.141	-0.138	-0.151	-0.152	-0.014
$\rho_{\alpha\gamma}$	-0.15/0	-0.170	-0.127	-0.165	-0.159	-0.013
$\rho_{\beta\gamma}$	0.25/0	0.302	0.239	0.264	0.233	-0.001

Table A2: Simulation results from 100 replications. Mean standard errors and standard deviation of parameter estimates across replicates.

	A $n = 1000$ $T = 5$ $\rho \neq 0$		B $n = 1000$ $T = 10$ $\rho \neq 0$		C $n = 1000$ $T = 15$ $\rho \neq 0$		D $n = 1500$ $T = 10$ $\rho \neq 0$		E $n = 1000$ $T = 10$ $\rho = 0$	
Parameter	Mean SE	SD	Mean SE	SD	Mean SE	SD	Mean SE	SD	Mean SE	SD
Coefficients										
α_0	0.159	0.148	0.119	0.096	0.103	0.108	0.094	0.102	0.117	0.105
α_1	0.104	0.091	0.088	0.088	0.085	0.088	0.074	0.068	0.091	0.082
β_0	0.156	0.158	0.106	0.111	0.096	0.093	0.085	0.071	0.112	0.128
β_1	0.102	0.093	0.073	0.069	0.070	0.059	0.061	0.058	0.079	0.068
γ_0	0.117	0.110	0.072	0.076	0.064	0.056	0.056	0.059	0.077	0.076
γ_1	0.052	0.051	0.037	0.036	0.036	0.031	0.030	0.030	0.040	0.041
Random effect parameters										
σ^2_α	0.658	0.627	0.461	0.424	0.389	0.423	0.365	0.335	0.454	0.484
σ^2_β	0.419	0.438	0.243	0.229	0.205	0.193	0.192	0.194	0.254	0.247
σ^2_γ	0.068	0.066	0.045	0.039	0.039	0.041	0.033	0.032	0.045	0.051
$\rho_{\alpha\beta}$	0.166	0.171	0.109	0.111	0.097	0.088	0.090	0.084	0.117	0.125
$\rho_{\alpha\gamma}$	0.215	0.178	0.135	0.125	0.118	0.104	0.108	0.112	0.141	0.130
$\rho_{\beta\gamma}$	0.194	0.173	0.130	0.121	0.116	0.103	0.100	0.102	0.145	0.147

References

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*, 172(3):579–598.