When “contexts” are geographical areas, is multilevel model still a good choice to model hierarchical data, or a new approach is needed?

Guanpeng Dong
Sheffield Methods Institute, The University of Sheffield
Email: guanpeng.dong@sheffield.ac.uk

The University Of Sheffield.

Contents

- Illustration of different dependence effects in MLM
- Integrated spatial and multilevel modelling
- Some applications
- A brief summary

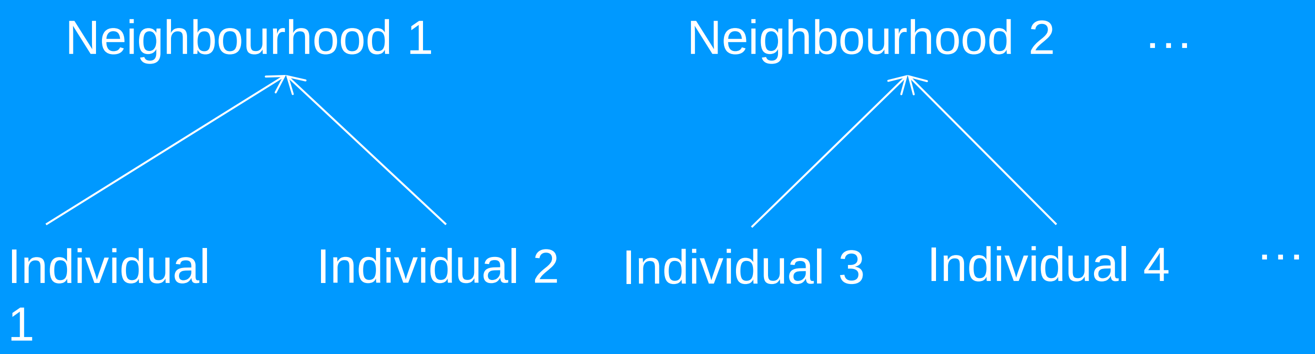
Sheffield Methods Institute.

Types of dependence effect in MLM

A simple vertical hierarchy

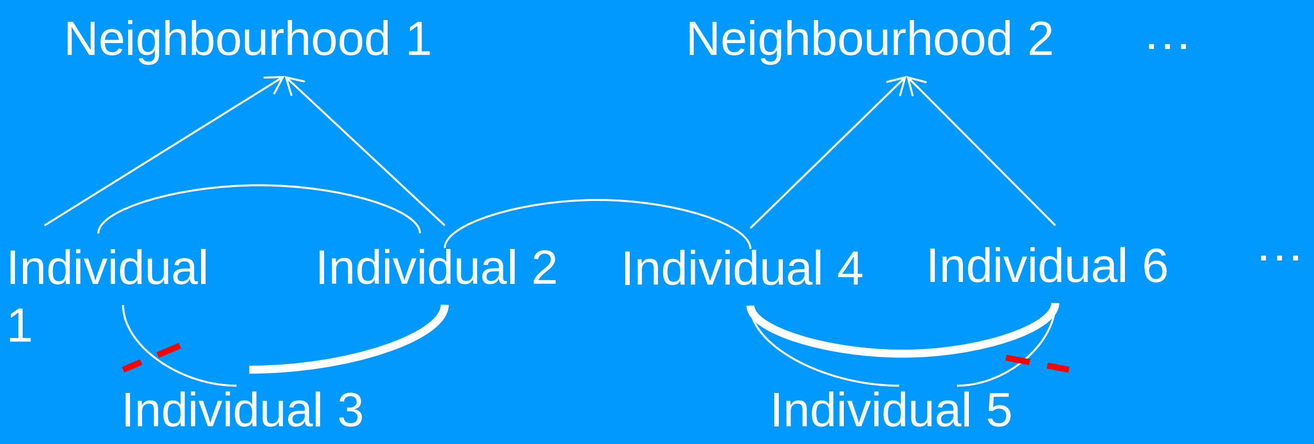
		Schools
		↑
Neighbourhoods	Neighbourhoods	Neighbourhoods
↑	↑	↑
Individuals	Houses	Pupils

A vertical effect



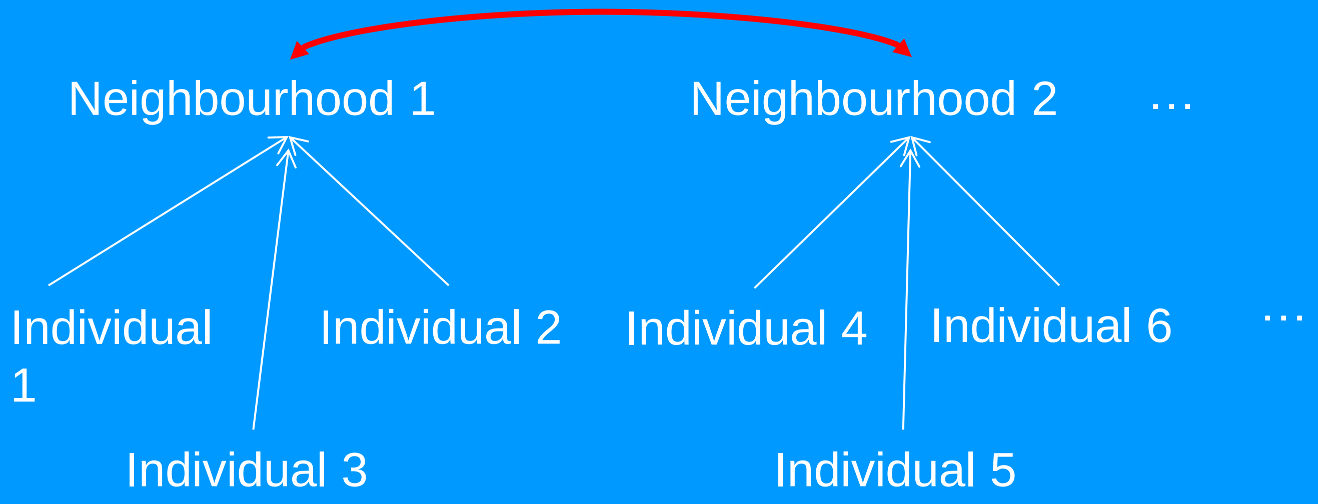
Correlations of individuals amongst the same neighbourhood; The Variance Partition Coefficient (VPC)—the proportion of the neighbourhood-level variance in the total variance

An interaction or horizontal effect (I)



Correlations of individuals might not be simply the VPC; Correlations or interactions between individuals amongst different neighbourhoods

An interaction or horizontal effect (II)



<u>A horizontal interaction effect between neighbourhoods</u>

Vertical and horizontal effects

- Vertical effect—for instance, school effects, neighbourhoods… is one of the most important motivations of the use of MLM
- Horizontal effect—interactions/interdependence/spillovers… is the reason of using spatial statistics/econometric models

Different views of space (or spatial effect)

- MLM—
  - a more discrete view of space and spatial effects. E.g. neighbourhoods are interchangeable without considerations to how neighbourhoods are arranged over space
  - Intra-space (intra-neighbourhood dependency)
  - Multi-scale modelling (individuals and neighbourhoods)

Different views of space (or spatial effect)

- Spatial statistics and econometrics—
  - a more continuous view of spatial effect. E.g. how neighbourhoods are spatially arranged is essential.
  - Inter-space (inter-neighbourhood) dependency
  - Single-scale modelling (either individual-level modelling or neighbourhood-level modelling, but not both)

A simple random intercept MLM

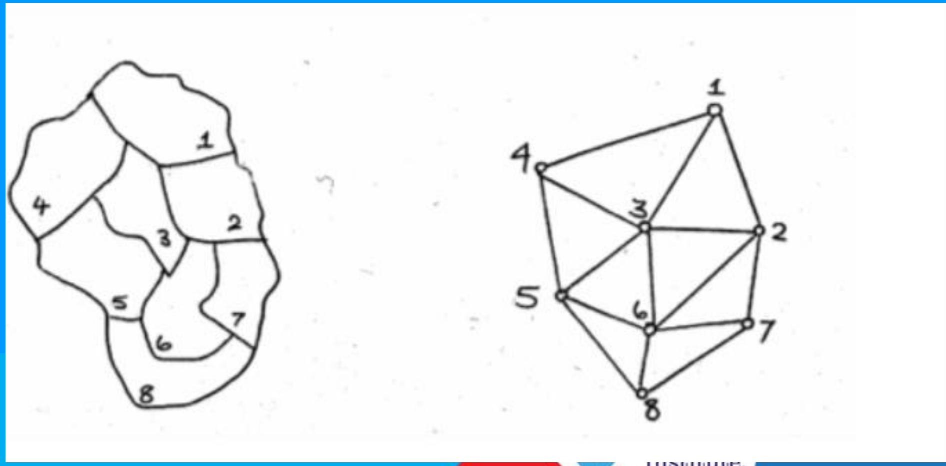
$$y_{ij} = \boldsymbol{\beta}_0 + x_{ij}^T \boldsymbol{\beta} + x_j^T \gamma + u_j + \varepsilon_{ij}$$
$$\mathrm{var}(\varepsilon_{ij}) = \sigma_e^2; \mathrm{var}(u_j) = \sigma_u^2$$
$$\mathrm{cov}(\varepsilon_{ij}, u_j) = 0; \mathrm{cov}(\varepsilon_{ij}\varepsilon_{i'j'}) = 0 \text{ if } j \neq j'$$

- No horizontal effects are posited.
- Only the vertical effect is modelled.

The most popular spatial econometric model—a spatial simultaneous autoregressive (SAR) model

$$Y = \rho WY + X\beta + \varepsilon \quad \varepsilon \sim N\left(0, I_N\sigma^2\right)$$

- *W* is the spatial weights matrix specifying how places (neighbourhoods) are connected *a prior*, sharing boundaries, for instance.



	1	2	3	4	5	6	7	8
1	0	1	1	1	0	0	0	0
2	1	0	1	0	0	1	1	0
3	1	1	0	1	1	1	0	0
4	1	0	1	0	1	0	0	0
5	0	0	1	1	0	1	0	1
6	0	1	1	0	1	0	1	1
7	0	1	0	0	0	1	0	1
8	0	0	0	0	1	1	1	0

- Useful to capture interactions amongst individuals or neighbourhoods, but…
  - What if there are geographical differences around the global model?
  - What if there are interaction effects operating at different scales?

Conceptually

- Vertical effects—the effects upon individuals from geographical contexts
- Horizontal effects—the effect acting between individuals and/or geographical contexts. E.g.
  - House price modelling, both the effect from nearby properties and the effect from the immediate neighbourhood and nearby neighbourhoods
  - Imagine, for example, a survey of attitudes/behaviours of residents living in communities; or

Risk of confounding

- A classic MLM might treat an interaction (horizontal) effect as a contextual (vertical) effect
- A classic spatial econometric model might treat a contextual (vertical) effect as an interaction (horizontal) effect
- The use of different models tend to claim findings of different effects

The way forward?

- An integrated spatial and multilevel modelling framework
  - A hierarchical spatial autoregressive model—a hybrid approach of spatial econometrics and multilevel modelling
  - A spatial random slope multilevel model—a hybrid approach of spatial statistics and multilevel modelling
  - Others?

A hierarchical spatial autoregressive model

Dong and Harris (2015). Spatial Autoregressive Models for Geographically Hierarchical Data Structures. *Geographical Analysis*, 47, 173–191

Model specification

Hierarchical spatial autoregressive model (HSAR)

$$\boxed{y = \rho W y} + X\beta + Z\gamma + \boxed{\Delta\boldsymbol{\theta}} + \varepsilon$$

Vertical effects

$$\boxed{\boldsymbol{\theta} = \lambda M \boldsymbol{\theta} + u}$$

Horizontal effects at two levels

$$\Delta = \begin{bmatrix} l_1 & 0 & \cdots & 0 \\ 0 & l_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & l_J \end{bmatrix}$$

Δ is a N by J block diagonal random effect design matrix;
$\boldsymbol{\theta}$, is the unobserved contextual/neighbourhood effects.
A SAR process is imposed to $\boldsymbol{\theta}$ to control the spatial dependence at this level.
ρ and λ measures the intensity of (spatial) interactions/correlations.
W (N by N) and M (J by J) are two spatial weights matrices at the two levels.

Model estimation

A Bayesian MCMC approach (please refer to Dong and Harris (2015) for detailed MCMC algorithms)**. Very briefly:**

$$P(\boldsymbol{\theta}^* \mid \text{Data}) \propto P(\text{Data} \mid \boldsymbol{\theta}^*) \times P(\boldsymbol{\theta}^*) \text{ where } \boldsymbol{\theta}^* = \{\rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_u^2\}, \text{Data} = \{Y, X, W, M\}$$

P (Data | $\boldsymbol{\theta}^*$) is the likelihood function of our model, we have seen it a lot of times in spatial econometrics textbooks

$$L(Y|\rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_u^2, \sigma_e^2) = (2\pi\sigma_e^2)^{-N/2}|A|\exp\{-(2\sigma_e^2)^{-1}(AY - X\boldsymbol{\beta} - \Delta\boldsymbol{\theta})'(AY - X\boldsymbol{\beta} - \Delta\boldsymbol{\theta})\}$$

Based on this and prior distributions for each unknown parameter, we get the posteriors,

$$P(\rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_u^2 \mid Y) \propto L(Y \mid \rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \sigma_u^2) \times P(\boldsymbol{\beta}) \times P(\rho) \times P(\lambda) \times P(\boldsymbol{\theta} \mid \lambda, \sigma_u^2) \times P(\sigma_u^2) \times P(\sigma_e^2)$$

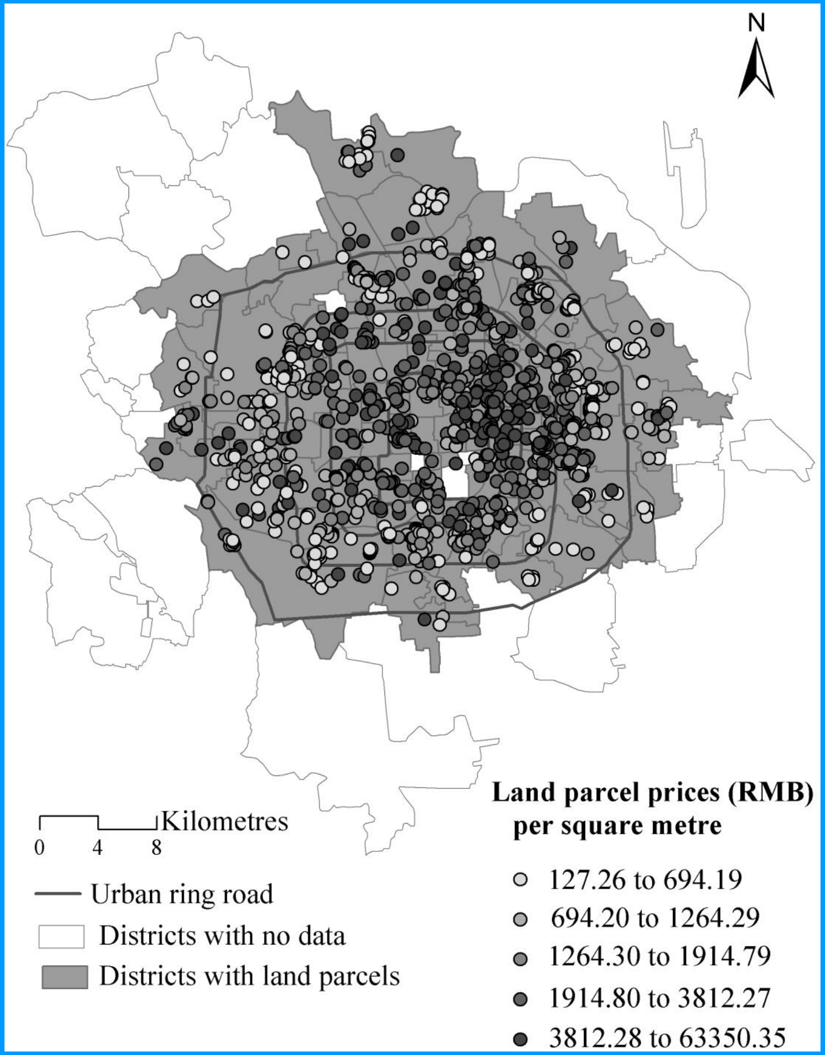
- The R code for implementing this model is provided as a supporting material in Dong et al. (2015). Multilevel Modelling with Spatial Interaction Effects. *PLOS ONE*, 10(6).

A simple assessment of MLM via Monte Carlo simulations

- The purpose
  - To assess how the estimation of MLM would be affected in various levels of horizontal effects
  - To test the practicality of our approach
  - Not a complete study to compare MLM and HSAR

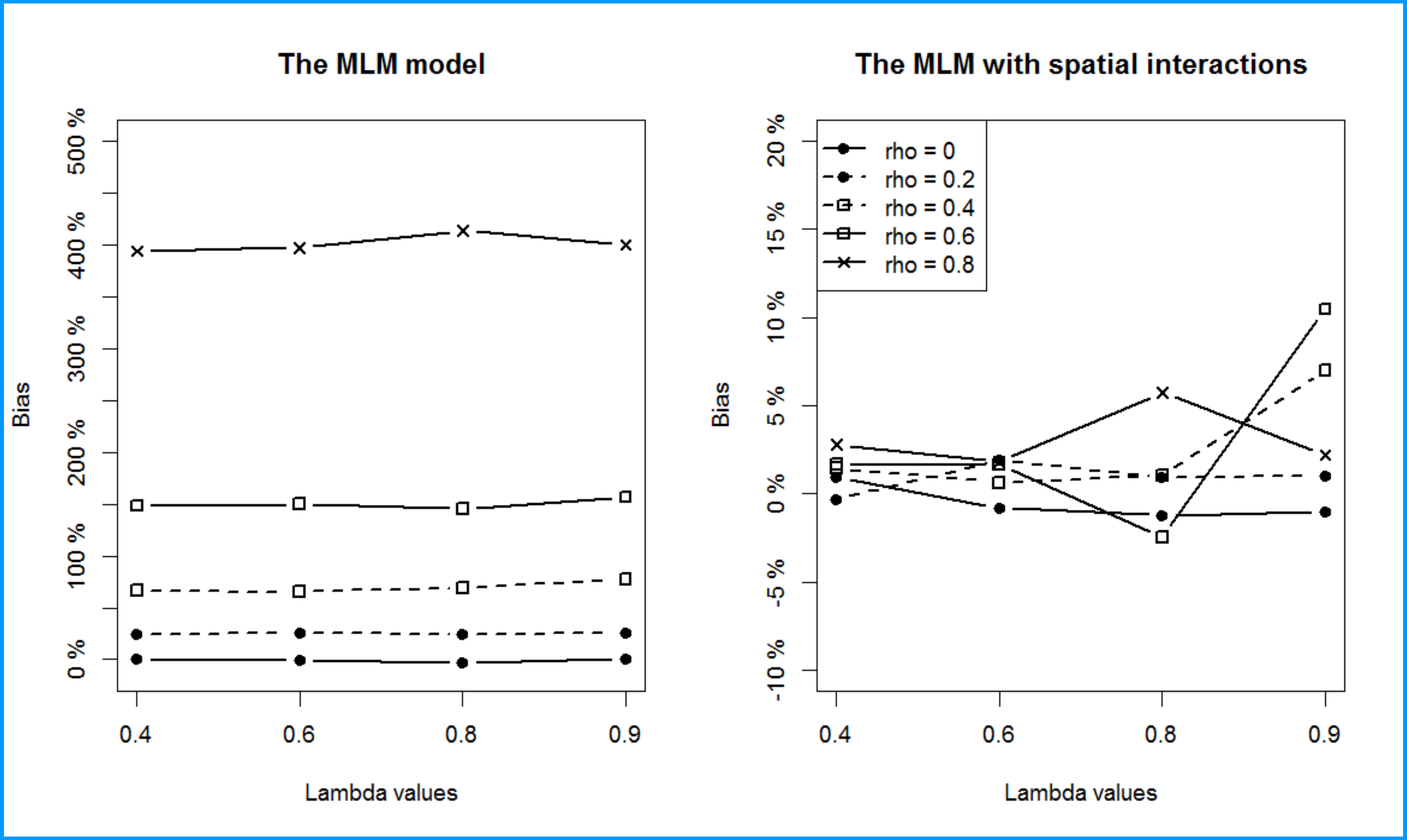
Data generating process

- Data generating process
  - Follow a HSAR DGP
  - Real-world geography of the residential land parcels data in Beijing, China. There are 1117 land parcels (lower level units) situated into 111 districts (higher level units), which has been explored by Harris, Dong and Zhang (2013).


N
Kilometres
0 4 8
Urban ring road
Districts with no data
Districts with land parcels
Land parcel prices (RMB) per square metre
127.26 to 694.19
694.20 to 1264.29
1264.30 to 1914.79
1914.80 to 3812.27
3812.28 to 63350.35

- Key model parameters
  - ρ—the land parcel-level interaction parameter. [0, 0.2, 0.4, 0.6, 0.8]
  - λ—the district/neighbourhood-level interaction parameter. [0.4, 0.6, 0.8, 0.9]
  - W is specified by a Gaussian kernel with a distance threshold of 1.5km.
  - M is based on the contiguity of districts.
- 200 simulated data samples are generated using HSAR as DGP; The same priors and hyerpriors are used for MLM and HSAR
- The bias and root-mean-square error (*RMSE*) of model parameters presented as percentage of their true values are used to assess model performance

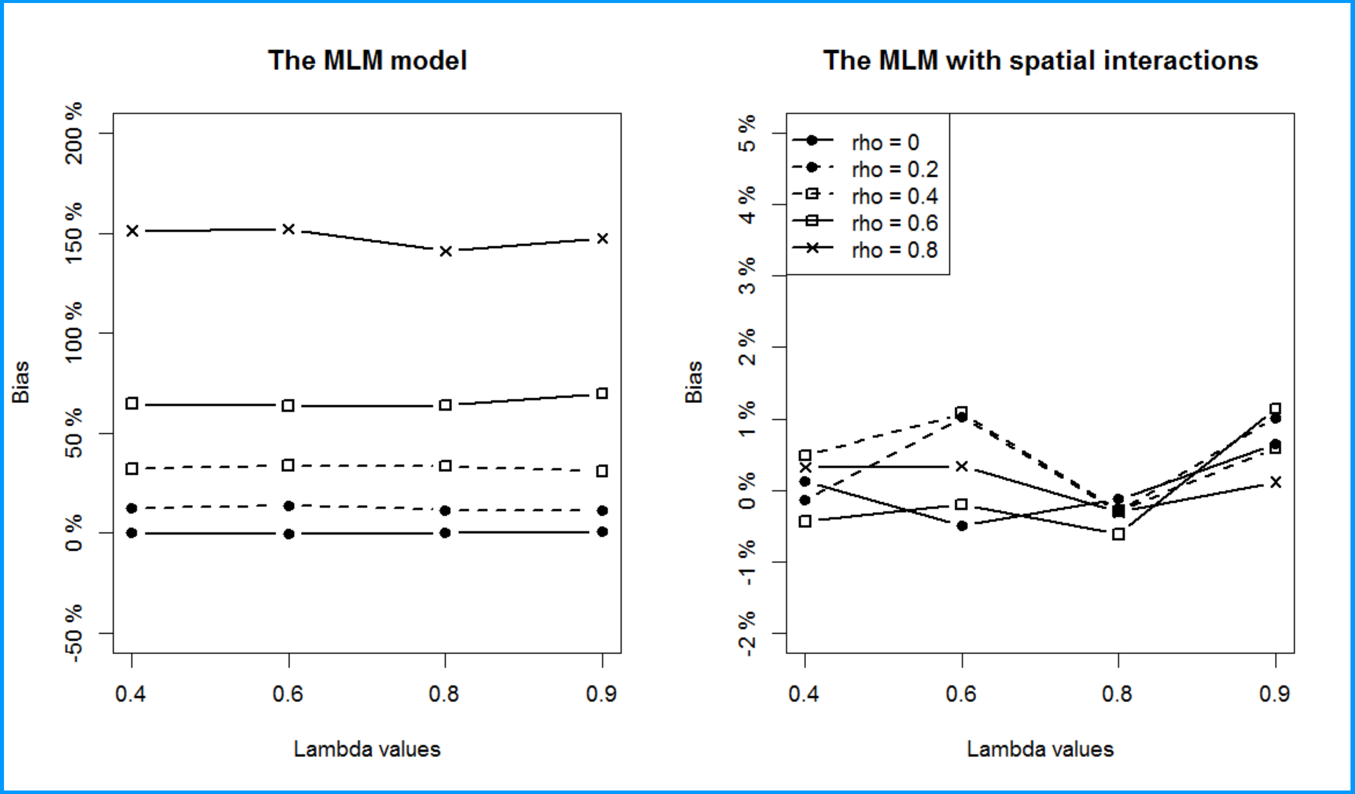
Simulation results for the Intercept term



(Note the different y-scales)

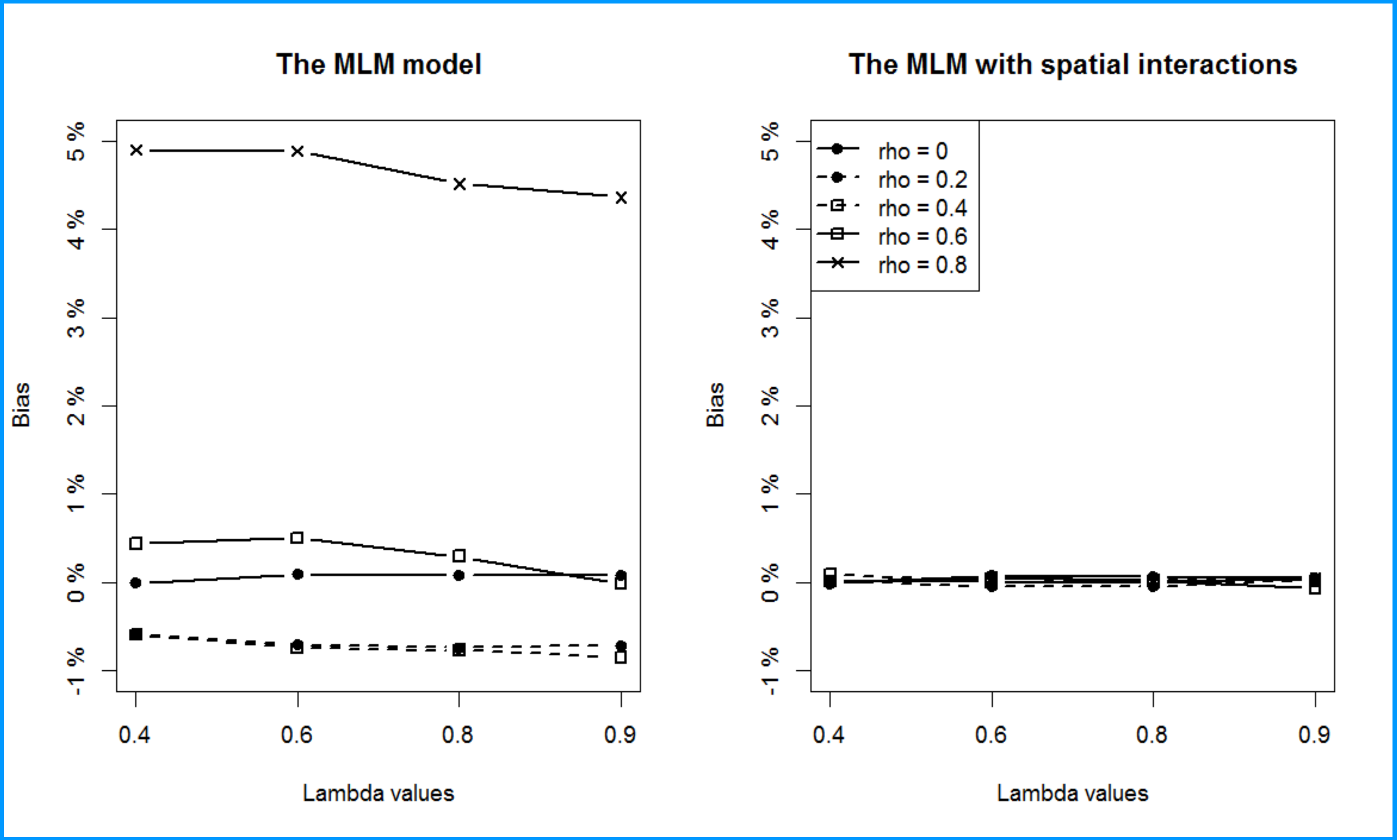
Details see Dong et al. (2015)

the higher-level variables (often referred to as contextual effects)

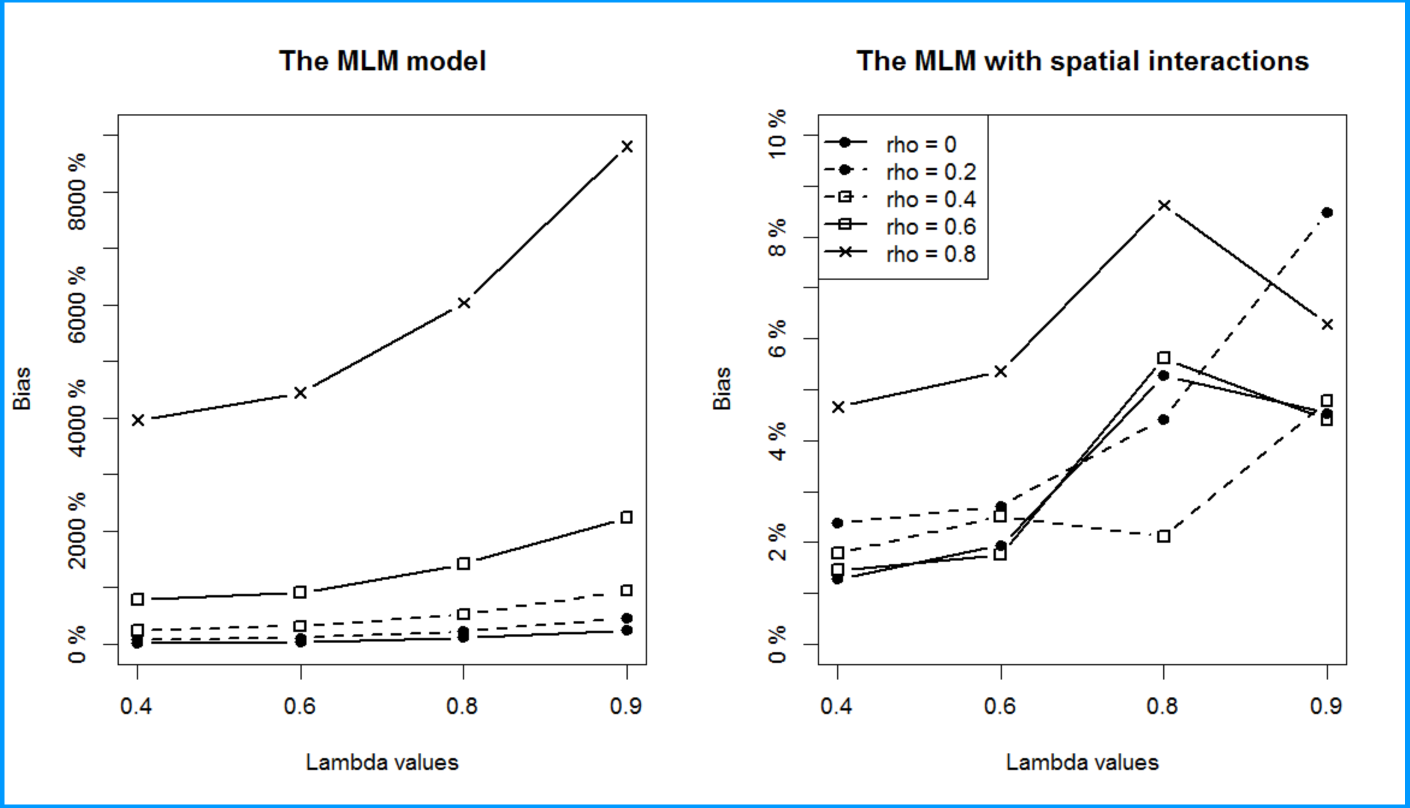


(Note the different y-scales)

the lower-level variables

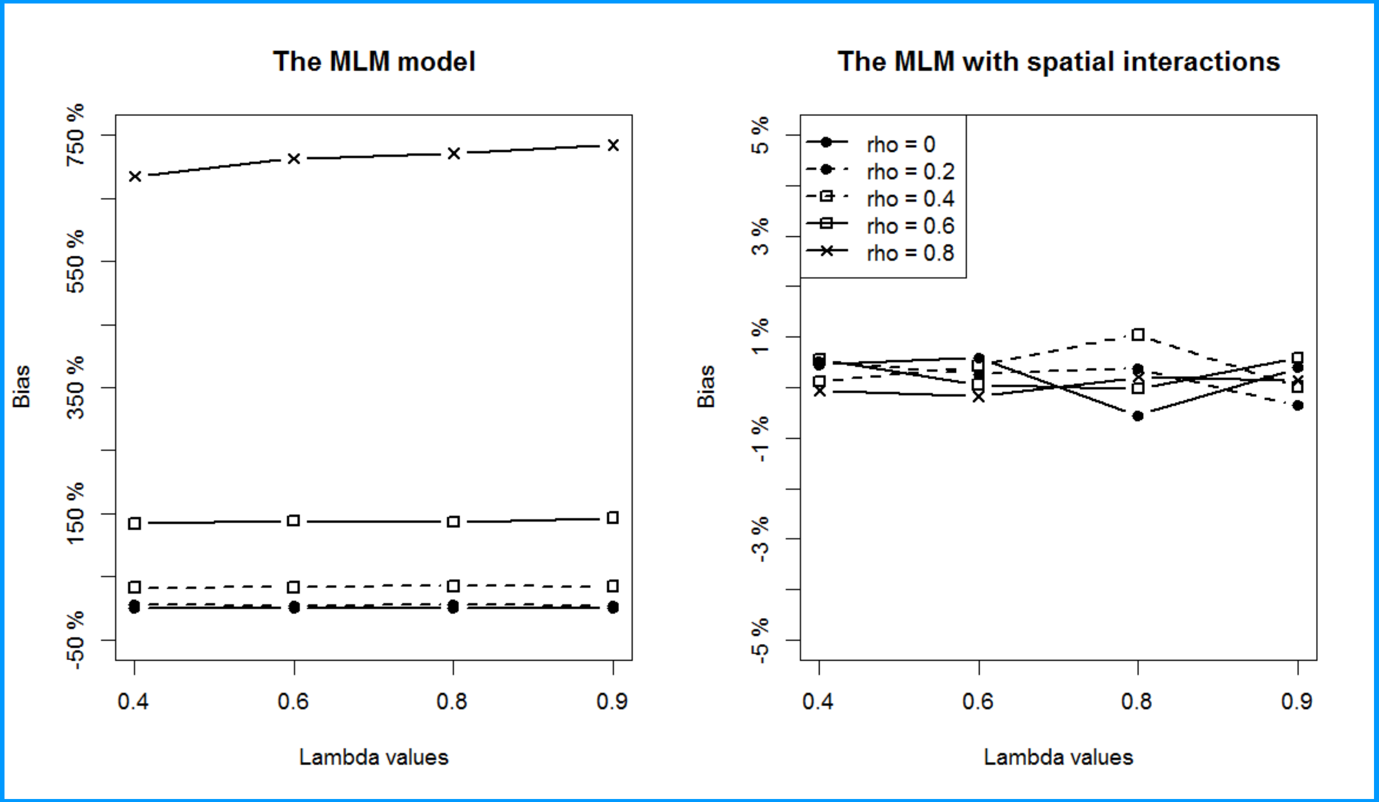


the higher-level variance



(Note the different y-scales)

the lower-level variance



(Note the different y-scales)

The University Of Sheffield.

Empirical evaluation --- does it matter in real-world data ?

	The HSAR				***The MLM***			
	Posterior mean	Std. Error	2.5%	97.5%	Posterior mean	Std. Error	2.5%	97.5%
Intercept	11.357*	0.958	9.448	13.248	13.627*	0.702	12.211	14.998
Logarea	-0.023	0.018	-0.059	0.013	-0.031	0.019	-0.068	0.006
LogDcbd	-0.362*	0.102	-0.576	-0.168	-0.373*	0.075	-0.520	-0.227
LogDsubway	-0.177*	0.042	-0.258	-0.097	-0.198*	0.042	-0.278	-0.115
LogDele	-0.015	0.039	-0.091	0.062	-0.054	0.038	-0.127	0.019
LogDpark	-0.148*	0.061	-0.266	-0.031	-0.245*	0.056	-0.359	-0.136
LogDriver	0.099*	0.036	0.031	0.168	0.137*	0.036	0.069	0.206
Popden	*0.019*	*0.013*	*-0.007*	*0.044*	*0.029**	*0.014*	*0.001*	*0.056*
Buildings1949	-1.082*	0.518	-2.067	-0.050	-1.380*	0.426	-2.211	-0.544
Crimerate	0.001	0.008	-0.015	0.017	0.010	0.010	-0.010	0.029
Year dummies		yes				yes		
ρ	0.174*	0.036	0.104	0.246	*NA*	*NA*	*NA*	*NA*
λ	0.760*	0.145	0.387	0.959	*NA*	*NA*	*NA*	*NA*
σ_u^2	$\boxed{0.052^*}$	0.021	0.020	0.102	$\boxed{0.125^*}$	0.030	0.075	0.192
σ_ε^2	0.574*	0.025	0.525	0.624	0.579*	0.026	0.530	0.633

Sheffield Methods Institute.

A brief summary

- It seems important to separate horizontal and vertical effects due to the risk of confounding
- Although the method is developed in the context of housing price modelling, the method should be able to be applied to social networks among individuals or friendship networks in pupils

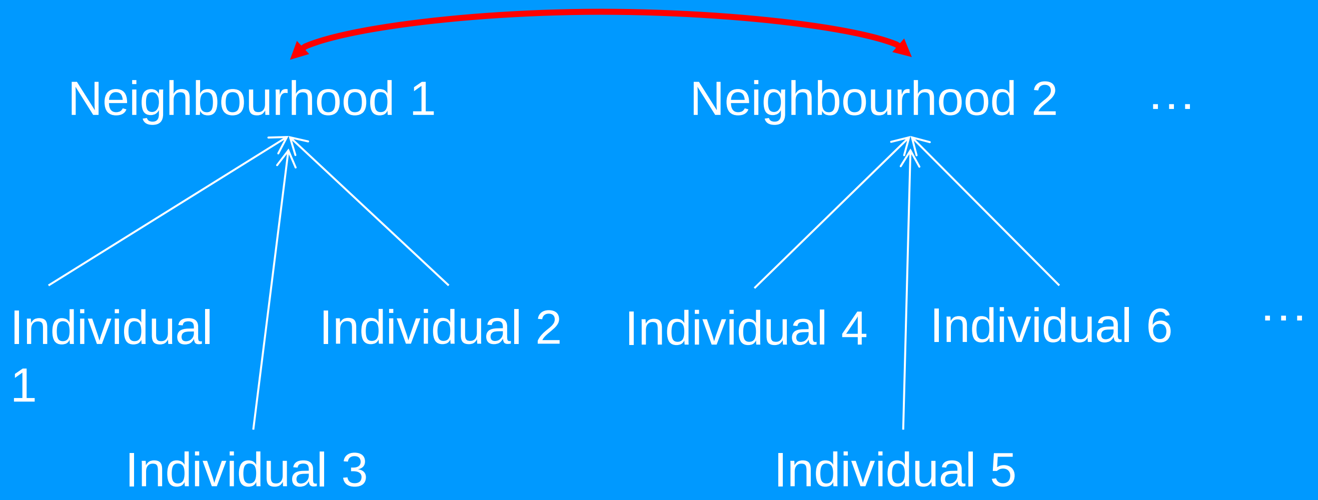
A spatial random slopes multilevel model—An extension of random slopes multilevel models with spatial effects

Dong et al. (2015, forthcoming). Spatial Random Slope Multilevel Modelling using Multivariate Conditional Autoregressive Processes. *Annals of the Association of American Geographers*.

The University Of Sheffield.

Sheffield Methods Institute.

The data structure under study



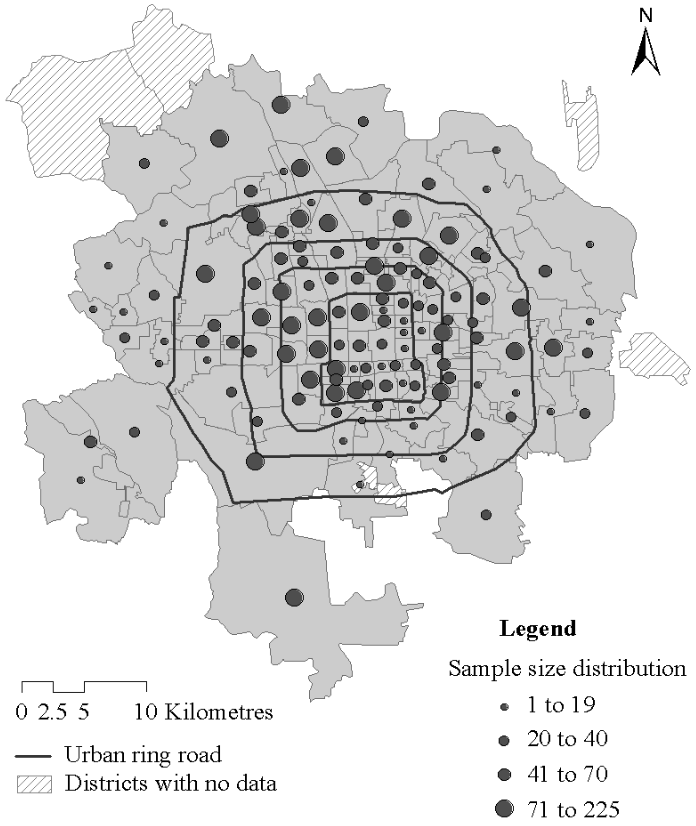
A horizontal interaction effect between neighbourhoods;
Random regression coefficients for individual-level covariates

The motivation

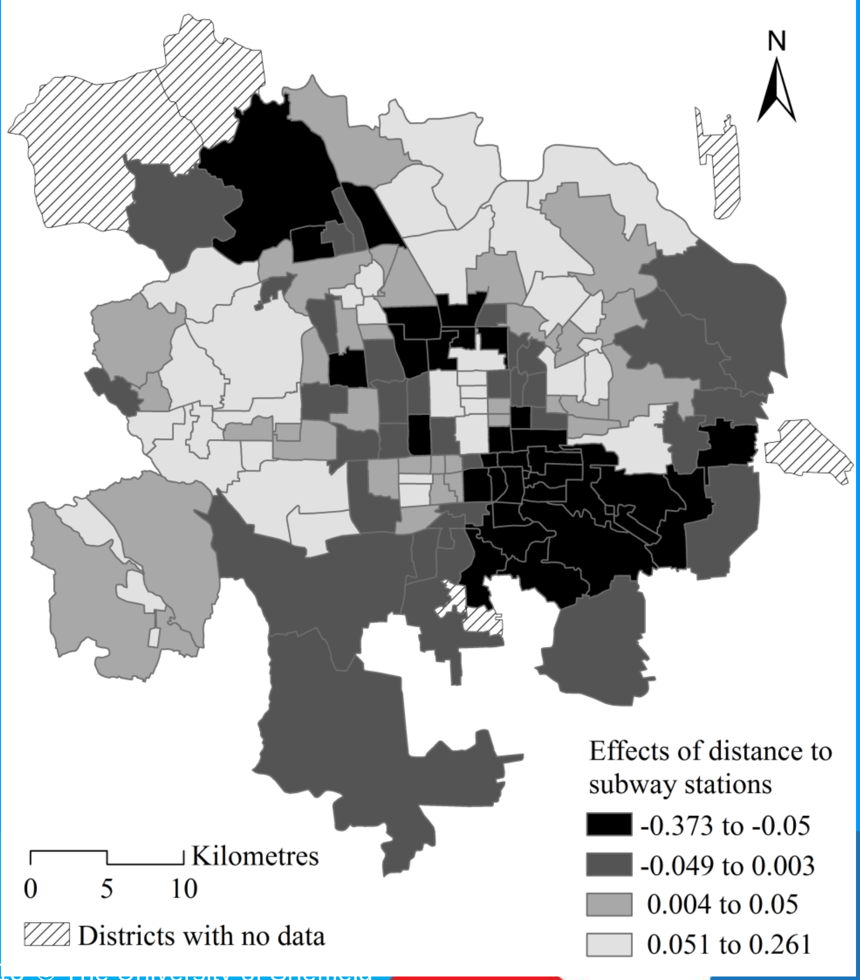
- Do locational factors such as the proximity to subway stations and green parks influence residents' travel satisfaction in Beijing?
- Does this effect vary across places?
- Is the variation purely random or in a way that is spatially structured due to the horizontal effect?

what is the data structure

- A large-scale household satisfaction survey in 2005 collected from Beijing, China. The effective sample includes 6467 individuals in 134 districts

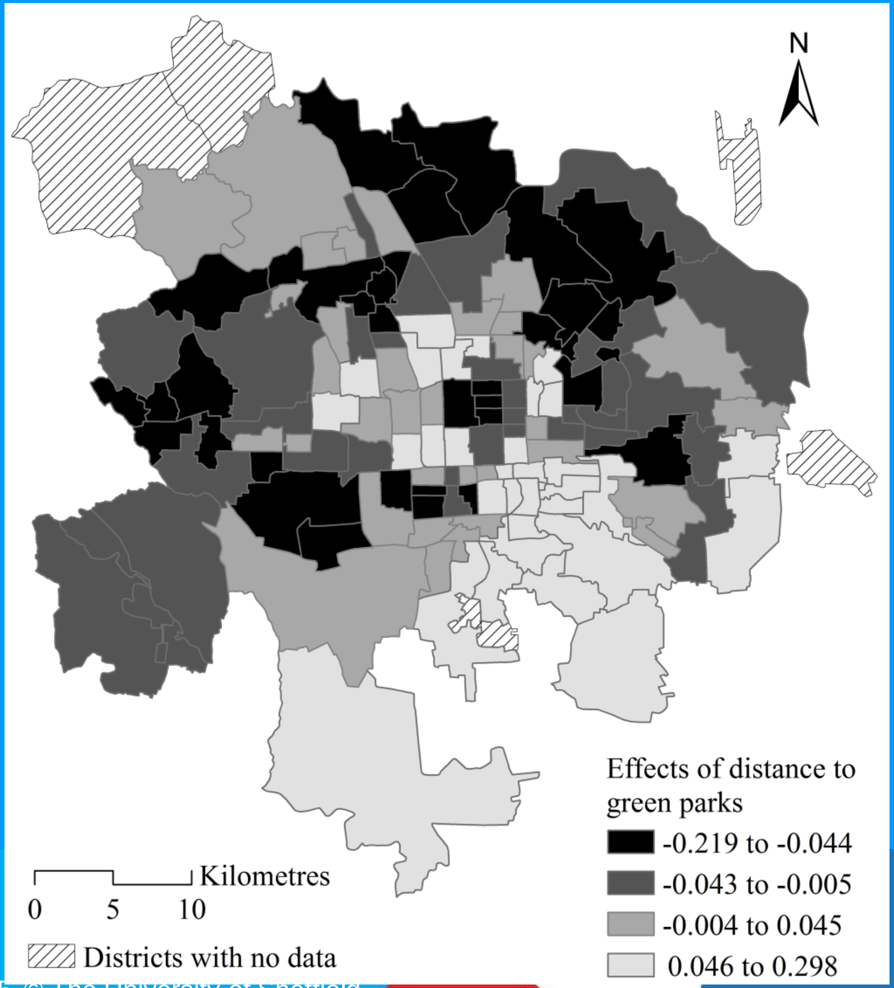


Initial random slope estimation from MLM



The Moran'I statistic for the random effects of the proximity to subway stations is 0.144 with a p-value equal to <0.001.

Not randomly distributed although it is assumed so in a random slopes multilevel model.



The Moran'I statistic for the random effects of the proximity to green parks is 0.127 with a p-value equal to <0.001.

Again, Not randomly distributed although it is assumed so in a random slopes multilevel model.

A random slope multilevel model

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + x_j^T\gamma_0 + \boxed{u_{0j}}$$

$$\beta_{1j} = \beta_1 + x_j^T\gamma_1 + \boxed{u_{1j}}$$

Higher level random effects

$$\mathrm{var}(\varepsilon_{ij}) = \sigma_e^2;\ \mathrm{cov}(u_{0j}, u_{1j}) = V_u = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01}^2 \\ \sigma_{u01}^2 & \sigma_{u1}^2 \end{bmatrix}$$

- No horizontal interaction/correlation effects are modelled.

Choices of modelling spatial dependence in this context

- By context,
  - Not only the intercept term is modelled as random effects, but also some regression coefficients
  - There might also be correlations between different random effect (*V*)
- SAR (Spatial econometric approach) is difficult to extended to a multivariate SAR, if possible
- Multivariate conditional autoregressive (CAR) process fits naturally here.

A brief illustration of CAR priors

$$\mathrm{E}(b_j|b_{-j}) = \frac{1}{w_{j+}} \sum_{k \sim j} b_k, \quad \mathrm{Prec}(b_j|b_{-j}) = \tau w_{j+}, \quad j = 1, \ldots, J,$$

- *b*, a $J \times 1$ random effect vector; the neighbourhood structure is defined by a simple contiguity based spatial weights matrix $W = (w_{jk} = 1 \text{ or } 0)$; $D_W = \text{diag}(w_{j+})$ where w_{j+} is the number of neighbours for region *j*
- τ is the precision parameter which is equal to $1/\sigma^2$
- The results in a unique Gaussian Markov Random Field (GMRF),

$b \sim \mathrm{MVN}(\mathbf{0}, \Omega_{iCAR})$ where $\Omega_{iCAR} = \tau(D_W - W)$

Details see, Besag et al. (1991); Banerjee et al. (2004); Rue and Held (2005)

A recently favoured Leroux et al. (1999) CAR (LCAR)

$$\mathrm{E}(b_j|b_{-j}) = \frac{\lambda}{1-\lambda+\lambda w_{j+}} \sum_{k\sim j} b_k, \ \mathrm{Prec}(b_j|b_{-j}) = \tau(1-\lambda+\lambda w_{j+})$$

- λ is spatial correlation parameter, measuring the intensity of spatial dependence/correlation; when λ -> 0, $b \sim N(0,1/\tau)$, an exchangeable prior; when λ -> 1, $b \sim iCAR\ prior$
- The results in a unique GMRF, *b* ~ MVN(**0**, Ω_{LCAR}) where $\Omega_{\mathrm{LCAR}} = \tau^2(L_W - W), L_W = \mathrm{diag}\,(1-\lambda+\lambda w_{k+})$
- The advantages of LCAR over other CAR priors are extensively discussed via simulations in Lee (2011) and MacNab (2011).

A Multivariate LCAR

- For a multivariate LCAR prior, despite of its complexity, it has a concise distribution form (Gelfand and Vounatsou 2003),

$$b_{J,P} \sim MVN(0, \Omega_{MLCAR}) \text{ with } \Omega_{MLCAR} = (L_W - W) \otimes \Gamma$$

Γ is a P by P positive definite precision matrix, the inverse of which measures the (conditional) variance and covariance of different sets of random effects.

The University Of Sheffield.

A spatial random slopes multilevel model—Model specification

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + x_j^T \gamma_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + x_j^T \gamma_1 + u_{1j}$$

$$\boldsymbol{u} \sim MlCAR, cov(\boldsymbol{u}) = \boxed{(L - \lambda W)} \otimes \boxed{V_u^{-1}}$$

Spatial interactions across space

Correlations between random effects within space

Sheffield Methods Institute.

Software packages

- R-INLA (http://www.r-inla.org/)?
  - A R package for fast implementation of approximate Bayesian inference using integrated nested Laplace approximations.
  - iCAR and convolution CAR (BYM) priors are directly available while LCAR can be specified through a user specified precision matrix structure (generic 1)
  - However, there is no multivariate CAR priors available.
- Winbugs has a built-in multivariate iCAR, but not LCAR
- Others?

Model estimation

- Bayesian MCMC using a Gibbs sampler with Metropolis updates when required
- Regression coefficients, random effects, individual-level variance, and the precision matrix are updated using Gibbs sampler while the spatial correlation parameter (λ) is updated using an adaptive Metropolis step. Details are provided in Dong et al. (2015)

The full conditional posterior distribution of β

$$P(\boldsymbol{\beta} \mid Y, \boldsymbol{\theta}, \lambda, \sigma_e^2, \Gamma) \sim N(M_{\beta}, \Sigma_{\beta}),$$

$$\Sigma_{\beta} = \left[(\sigma_e^2)^{-1} X'X + T_0^{-1}\right]^{-1}; \; M_{\beta} = \Sigma_{\beta} \left[(\sigma_e^2)^{-1} X'(Y - Z\boldsymbol{\theta}) + T_0^{-1} M_{\mathbf{0}}\right]$$

The conditional posterior distributions for $\{\theta, \sigma_e^2, \Gamma\}$

$$P(\theta \mid Y, \boldsymbol{\beta}, \lambda, \sigma_e^2, \Gamma) \sim N(M_{\theta}, \Sigma_{\theta}),$$

$$\Sigma_{\theta} = \left[(\sigma_e^2)^{-1} Z'Z + (L_W - \lambda W) \otimes \Gamma\right]^{-1}; \; M_{\theta} = \Sigma_{\theta} \left[(\sigma_e^2)^{-1} Z'(Y - X\boldsymbol{\beta})\right].$$

$$P(\sigma_e^2 \mid Y, \boldsymbol{\beta}, \lambda, \theta, \Gamma) \sim IV(c_e, d_e)$$

$$c_e = N/2 + c_0; \; d_e = \frac{1}{2} \times (Y - X\boldsymbol{\beta} - Z\boldsymbol{\theta})'(Y - X\boldsymbol{\beta} - Z\boldsymbol{\theta}) + d_0$$

The conditional posterior distributions for $\{\theta, \sigma_e^2, \Gamma\}$

$P(\Gamma \mid Y, \boldsymbol{\beta}, \lambda, \sigma_e^2, \theta) \sim dwish(R^*, v^*)$,

$v^* = J + v_0$; $R^* = [R_0{}^{-1} + \boldsymbol{\theta}^{*\prime}(L_W - \lambda W)\boldsymbol{\theta}^*]^{-1}$. Where $\boldsymbol{\theta}^*$ is a J by p matrix with each column being the random effects that pertain to each individual level covariate.

The conditional posterior distribution for λ

$$P(\lambda \mid Y, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_e^2, \Gamma) \propto P(\boldsymbol{\theta} \mid \lambda, \Gamma) \times P(\lambda)$$
$$\propto |(L_W - \lambda W) \otimes \Gamma|^{1/2} \exp\{-1/2\boldsymbol{\theta}'[(L_W - \lambda W) \otimes \Gamma]\boldsymbol{\theta}\}$$

Therefore, an adaptive Metropolis step with an acceptance rate about 50% is used to update λ

The University Of Sheffield.

The advantage of the spatial random slope MLM against a standard MLM

- In our travel satisfaction study, the two models are compared using DIC (Spiegelhalter et al. 2002) and the pseudo-Bayes factor (PsBF).

	DIC	PsML	PsBF (in favour of MLM-MLCAR)
$\boxed{\text{MLM\_MLCAR}}$	17602.0	-8806.2	
MLM_MICAR	17608.7	-8809.2	
$\boxed{\text{non-spatial MLM}}$	17608.7	-8809.0	
MLM_MLCAR/MLM_MICAR	-6.7	3.0	20.6 (strong)
MLM_MICAR/ non-spatial MLM	-6.7	2.8	15.8 (strong)

Sheffield Methods Institute.

Discussion

- If correlations between different sets of random effects were not important or just not of our interest, model implementation would be much simpler, for example by using R-INLA package.

Discussion

- If correlations between different sets of random effects were not important or just not of our interest, model implementation would be much simpler, for example by using R-INLA package.
- A very brief demonstration of a spatial multilevel logistic model of environmental hazard effects on self-rated health

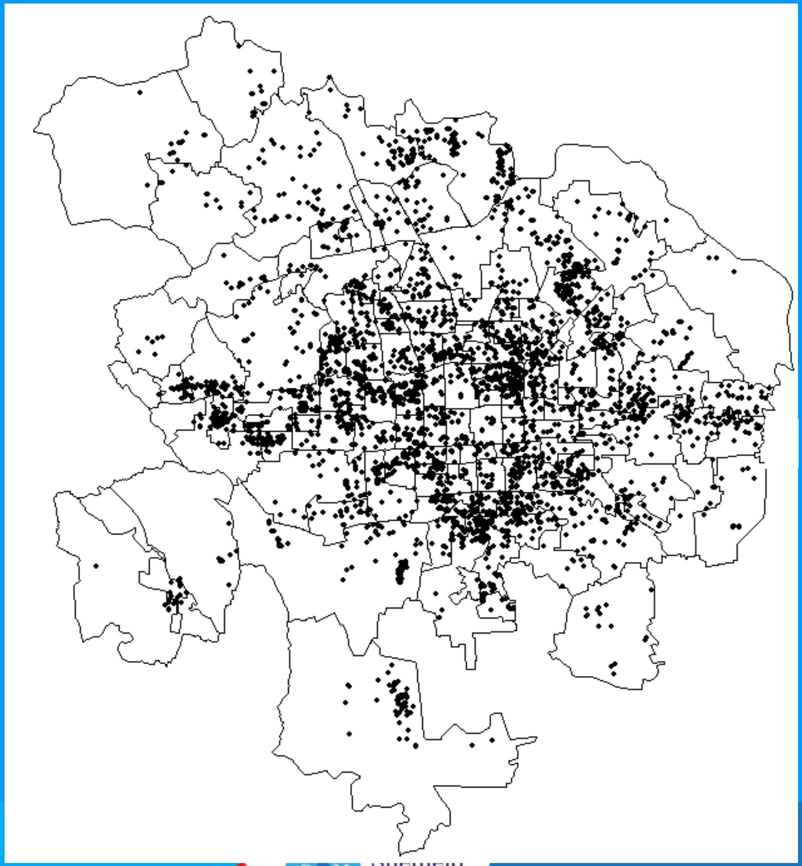
A spatial multilevel logistic model of environmental hazard effects on self-rated health

Dong et al. (2015, submitted). Perceived Environmental Hazards, Geography and Health in Beijing, China—A Bayesian Spatial Multilevel Logistic Regression Model.

The aim

- Whether geography matters in the self-rated health status in Beijing?
- How perceived environmental hazards influence self-rated health, controlling for …

Our data structure



The University Of Sheffield.

A spatial multilevel logistic regression model

$$\ln(p_{jk}/1 - p_{jk}) = \eta_{jk} = a + P'_{jk}\boldsymbol{\beta} + S'_{jk}\boldsymbol{\delta} + D'_{k}\boldsymbol{\varphi} + u_k$$

$$\boldsymbol{u} \sim MVN(\mathbf{0}, \Omega_{\mathrm{LCAR}}(\lambda, \tau^2)),$$

$$\{a, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\varphi}\} \sim N(0, b), \tau^2 \sim \mathrm{gamma}(e', f'), \mathrm{logit}(\lambda) \sim N(0,100)$$

Sheffield Methods Institute.

R-INLA implementation

- A formula object

formula.LCAR <- The outcome variable ~ Xβ +

f(Uid, model = "generic1",Cmatrix = district_mat_LCAR)

Or ***change*** the default hyperpriors

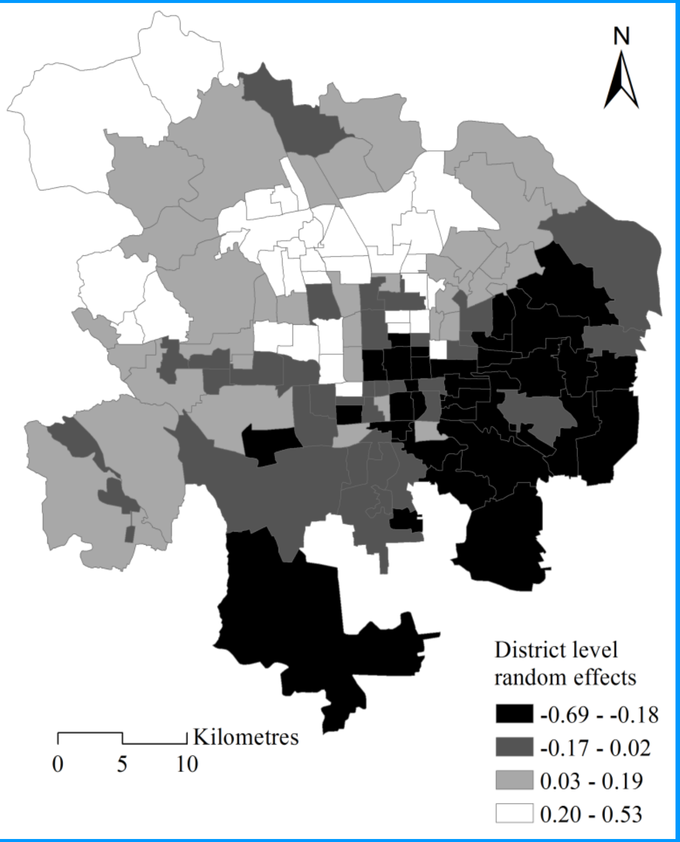
f(Uid,model = "generic1",Cmatrix = district_mat_LCAR,hyper=list(prec=list(prior="loggamma",param=c(*1,0.01*)),

beta=list(prior="logitbeta",param=c(*2,2*))))

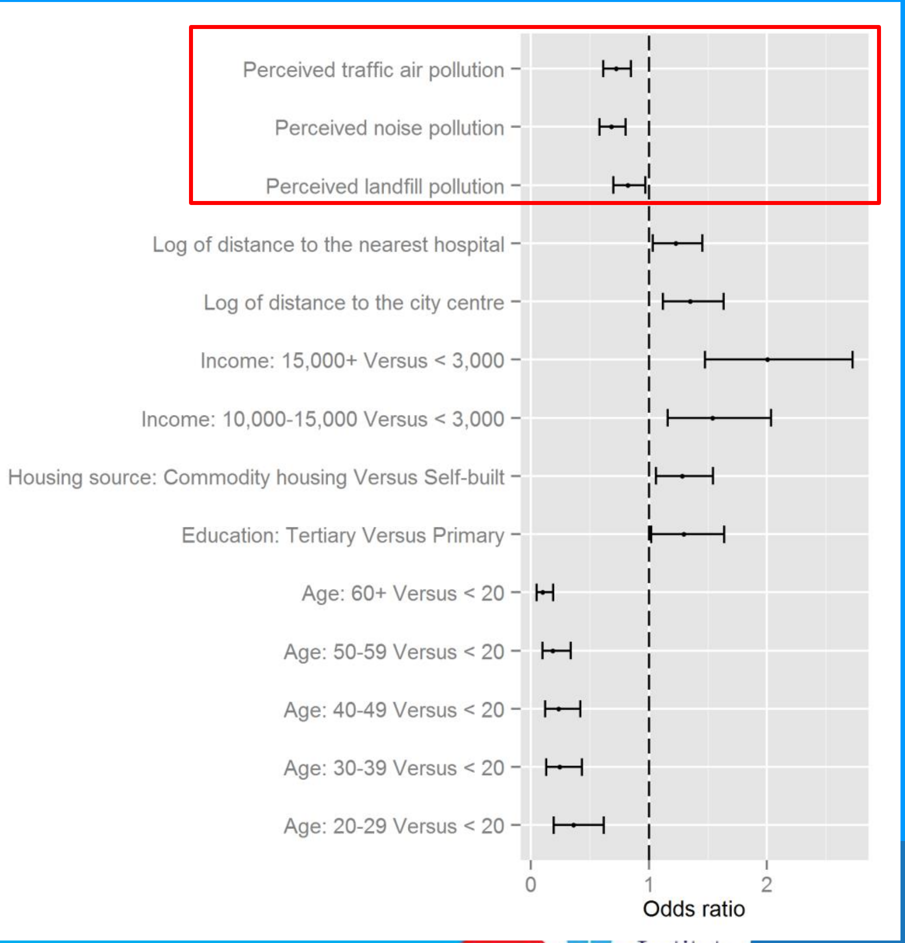
- Run the model

model.LCAR <- inla(formula.LCAR,data=model.data, family="binomial", control.predictor = list(compute = TRUE), control.compute=list(dic=TRUE,cpo=TRUE), control.inla=list(strategy="simplified.laplace"))

Results—the geography effect



Results—environmental hazards effect



A summary

- Geographical information is Not just a burden that usually complicates our analysis, but also an Opportunity to better understand how context influences our behaviours
- Detailed geographical information is increasingly available to researchers, such as social media and other valuable ‘big data’

- It is not necessary to include all types of horizontal and vertical effects in our research, just choose the one that you suspect might make a difference in terms of inference

Many thanks