Bayesian hierarchical models for small area data

with applications in social science

Guangquan Li Nicky Best Sylvia Richardson

Department of Epidemiology and Biostatistics
Imperial College London

NCRM and SRA joint event series in 2011 (May 11, 2011)

Outline

Outline

Small area data in the UK

- Administrative geography in UK includes
  - Census Output Area (COA; ∼300 people)
  - Electoral wards (∼ 500 to 2000 people)
  - Local authority districts (10's of thousands)
- Many administrative data sets contain information on social, economic and health indicators of the population.
  - Census: population and socio-economic indicators
  - Counts of births, deaths, cancer registrations, hospital admissions . . .
  - Counts of burglaries, car thefts . . .
  - Average income estimates from purposely conducted surveys.
  - etc.
- These data are often geo-referenced by small area
- Provide rich source of data for empirical social science research.

Main issue in the analysis of small area data

- "Small" does not necessarily refer to the geographical size of an area.
  - $\rightarrow$ A district in the UK, typically of several hundred thousand population, can be a "small area" if we are interested in some rare outcome (cancer, homicide).
- For our purposes, an area is considered to be small if
  1. outcome of interest is "rare", e.g., number of cancer deaths, burglaries and benefit claimants or
  2. the sample size is much smaller than the total population in the area, e.g., typically a survey only covers 1% of the population.
- Any direct measure based on the sparse data becomes unreliable

Main issue in the analysis of small area data

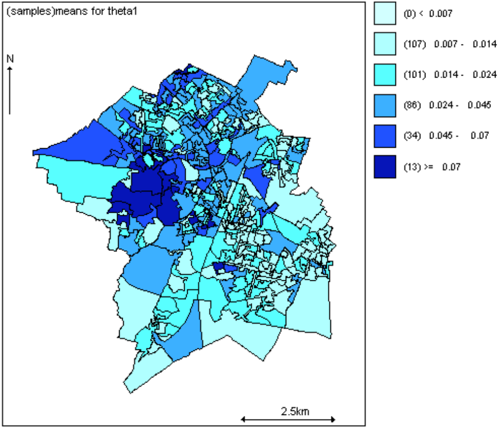
- "Small" does not necessarily refer to the geographical size of an area.
  - $\rightarrow$ A district in the UK, typically of several hundred thousand population, can be a "small area" if we are interested in some rare outcome (cancer, homicide).
- For our purposes, an area is considered to be small if
  1. outcome of interest is "rare", e.g., number of cancer deaths, burglaries and benefit claimants or
  2. the sample size is much smaller than the total population in the area, e.g., typically a survey only covers 1% of the population.
- Any direct measure based on the sparse data becomes unreliable $\rightarrow$ need more sophisticated statistical analysis techniques.

Bayesian hierarchical (multilevel) models

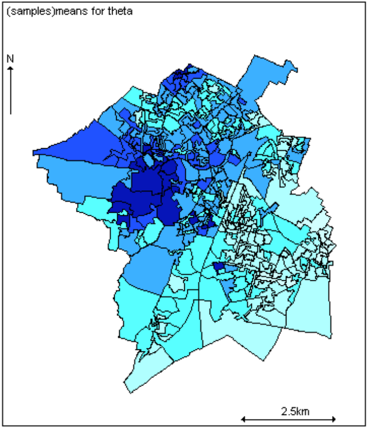
- Bayesian hierarchical models (BHM) offer a natural framework to combine information and hence to strengthen estimation.
- The sparsity issue in geo-referenced data can be addressed by "smoothing" over space or time or both.
- Idea is to "borrow information" from neighbouring areas or time periods to produce better (more stable, less noisy) estimates in each area.
- Modelling spatial or temporal structure achieved by appropriate choice of random effects distribution.
- Estimation and inference from these models are more easily achieved using a Bayesian model formulation + simulation-based estimation methods (e.g., MCMC).

Burglary rates in Cambridgeshire

Unsmoothed map



Smoothed by Bayesian models



Motivating examples

1. Small area estimation of income using Bayesian spatial models

2. Evaluation of the Cambridgeshire Police “No Cold Calling” scheme

Motivating examples

1. Small area estimation of income using Bayesian spatial models
    - Structure: geocoded survey data
    - Objective: to provide estimates of average income at some geographical level (e.g., MSOA)
    - BHM with spatially structured random effects shown to improve estimates for small areas with/without survey samples.
2. Evaluation of the Cambridgeshire Police “No Cold Calling” scheme

Motivating examples

1. Small area estimation of income using Bayesian spatial models
   - Structure: geocoded survey data
   - Objective: to provide estimates of average income at some geographical level (e.g., MSOA)
   - BHM with spatially structured random effects shown to improve estimates for small areas with/without survey samples.
2. Evaluation of the Cambridgeshire Police “No Cold Calling” scheme
   - Structure: time series data for a set of small areas
   - Objective: to evaluate impact of policy to reduce burglary rates in some areas
   - BHM smooths temporal trends and departures of trends in the NCC-targeted areas from the control trend can be better estimated using a hierarchical modelling structure.

Outline

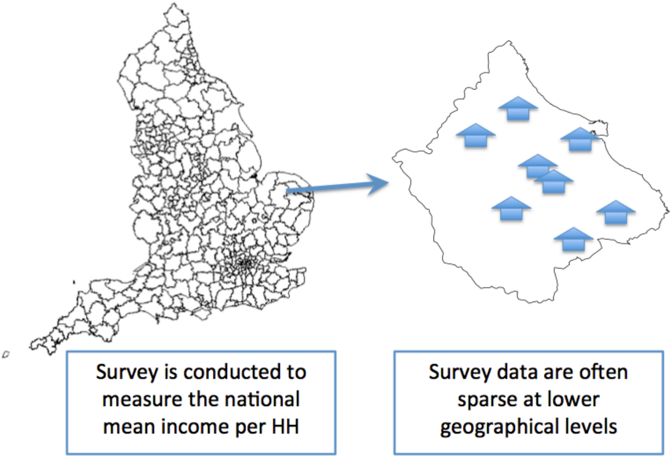
Income: Background

- National statistical offices are often required to provide statistical information about characteristics of the population, e.g., mean income, at several administrative or small area levels.
- Although reliable at regional or national level, survey data are often sparse when disaggregated to lower geographical areas.
  - Due to sampling design, some small areas may not even have survey samples.

Questions of interest are:

1. to provide reliable small area estimates for both the in-sample and off-sample areas;
2. to rank/classify areas to help inform policy making.

Income: Survey data at small area level



The empirical mean of the survey sample, $\hat{\bar{y}}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$, as an estimate of the mean income for small area i would be highly variable.

Income: BHM with spatial random effects

Let

- y_{ij} be the income of household j in area i that was included in the survey;
- $\mathbf{x}_{ij}$ be a vector of household level covariates

The following model is applied

$$\begin{aligned} y_{ij} &\sim \mathrm{N}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \alpha + \mathbf{x}_{ij} \cdot \beta + u_i + v_i \end{aligned}$$

where

- α and β are the regression coefficients and
- σ^2 is the sampling variability.

Income: Spatial smoothing

$$
\begin{aligned}
y_{ij} &\sim \mathrm{N}(\mu_{ij}, \sigma^2) \\
\mu_{ij} &= \alpha + \mathbf{x}_{ij} \cdot \beta + u_i + v_i
\end{aligned}
$$

- u_i and v_i are modelled as random effects.
- v_i are modelled as unstructured random effects, $v_i \sim \mathrm{N}(0, \sigma_v^2)$.

Income: Spatial smoothing

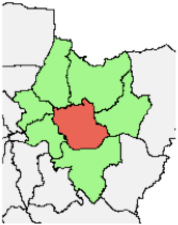
$$\begin{aligned} y_{ij} &\sim \mathrm{N}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \alpha + \mathbf{x}_{ij} \cdot \beta + u_i + v_i \end{aligned}$$

- u_i and v_i are modelled as random effects.
- v_i are modelled as unstructured random effects, $v_i \sim \mathrm{N}(0, \sigma_v^2)$.
- The (intrinsic) Conditional AutoRegressive (CAR) model is assigned to $\mathbf{u}$ ($\equiv u_{1:N}$) to allow for spatial dependence:

$$\begin{aligned} u_i | \mathbf{u}_{-i} &\sim \mathrm{N}(m_i, s_i) \\ m_i &= \sum_{k \in NB_i} u_k / n_i \\ s_i &= s / n_i \end{aligned}$$

m_i = average of u_k in neighbouring areas

s_i is random effect variance weighted by no. of neighbours



NB_i the set of neighbours for area i

n_i is number of neighbours

Income: Spatial smoothing

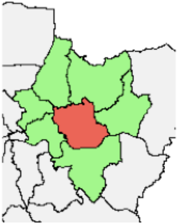
$$
\begin{aligned}
y_{ij} &\sim \mathrm{N}(\mu_{ij}, \sigma^2) \\
\mu_{ij} &= \alpha + \mathbf{x}_{ij} \cdot \beta + u_i + v_i
\end{aligned}
$$

- u_i and v_i are modelled as random effects.
- v_i are modelled as unstructured random effects, $v_i \sim \mathrm{N}(0, \sigma_v^2)$.
- The (intrinsic) Conditional AutoRegressive (CAR) model is assigned to $\mathbf{u}$ ($\equiv u_{1:N}$) to allow for spatial dependence:

$$
\begin{aligned}
u_i|\mathbf{u}_{-i} &\sim \mathrm{N}(m_i, s_i) \\
m_i &= \sum_{k \in NB_i} u_k / n_i \\
s_i &= s / n_i
\end{aligned}
$$

m_i = average of u_k in neighbouring areas

s_i is random effect variance weighted by no. of neighbours



NB_i the set of neighbours for area i

n_i is number of neighbours

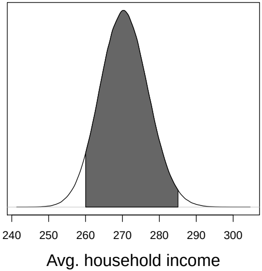
- The SAE for area i is given by $\hat{\mu}_i = \hat{\alpha} + \hat{\beta} \cdot \overline{\mathbf{X}}_i + \hat{u}_i + \hat{v}_i$.

Income: Benefits of spatial BHM

- Spatial structure allows us to borrow information from neighbouring areas when estimating the spatially-correlated random effects, which has been shown to improve SAE, in particular, for areas without survey samples.

- Being Bayesian, probability statements about the parameters can be easily made, i.e.,

$$P(b_L < \text{Avg. Income} < b_U)$$



- Ranking and classification of areas, a common problem faced by statistical bureaus, can be produced easily. Uncertainty of the ranks, a crucial information to present, is also readily available.

Income: Average equivalised income per household in Sweden

- Data
  - Using the LOUISE Population Register in Sweden, 100 "mock" surveys were constructed.
  - 284 municipalities in Sweden in 1992.
  - Sample size: 0.1% of the total population (a total sample size of 3358 households) with some municipalities without survey data
  - True area values are known (so can be used for model evaluation)
  - Covariates: number of persons in HH, gender, age, education of the head of the HH.
- Models compared
  - With/without the random effects $u_i + v_i$.

Income: Improved SAE for in/off-sample areas

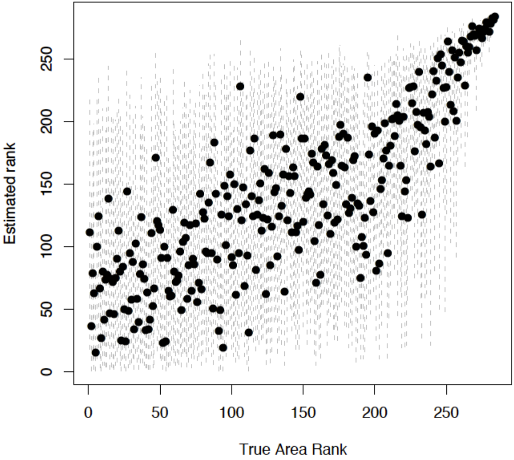
0.1% sample	In-sample areas		Off-sample areas	
	MARB	MRRMSE	MARB	MRRMSE
Empirical mean	0.015	0.176	-	-
Without $u_i + v_i$	0.058	0.064	0.069	0.074
With $u_i + v_i$	0.020	0.038	0.030	0.041

- **MARB**: Mean Absolute Relative Bias, measuring the bias of the estimator.
- **MRRMSE**: Mean Relative Root Mean Square Error, measuring the variability of the estimator.

1. Model-based methods can (a) reduce variability of SAE and (b) provide SAE for off-sample areas.
2. For both the in-/off-sample areas, the BHM with both spatially structured and unstructured random effects produces SAEs that are less biased and less variable than those from models without any random effects.

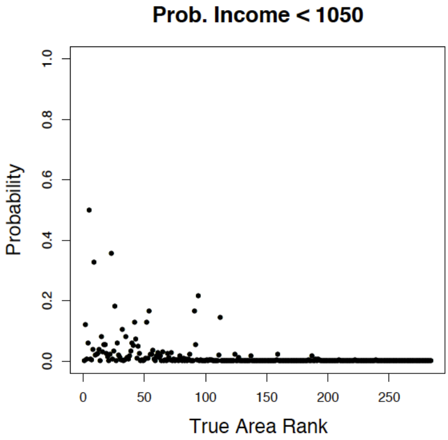
Income: Ranking and classification

- Ranking of areas can be produced by using point estimates (from either Bayesian or frequentist approach)
- but care must be taken in interpreting the results due to the considerable uncertainty in the ranks.



Income: Ranking and classification

- We suggest to rank areas using the posterior probability
  - e.g., rank the posterior prob. of household income less than £200.
- Broadly discriminate the rich from the poor where the latter may require appropriate assistance.

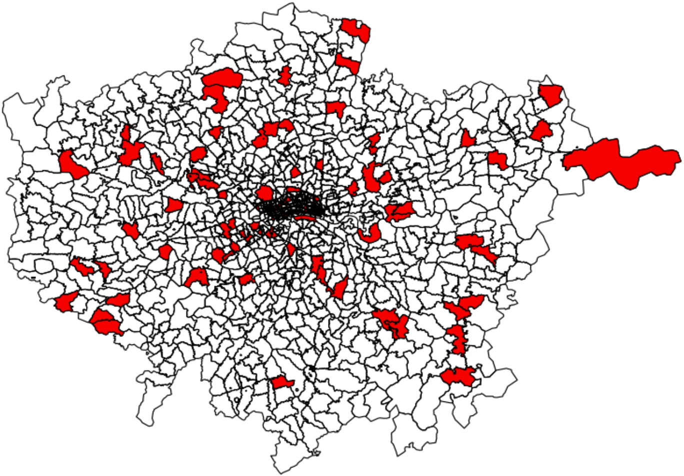


Income: Family Resources Survey (FRS) application (collaboration with ONS)

- Survey description
  - The FRS covers England and Wales.
  - The main interest is to provide estimates of average income per household at the Middle Super Output Area (MSOA) level.
  - The primary sampling unit is at the postcode sector (PCS) level.
- Main challenge
  - Roughly a third of a total 8569 PCS were sampled

    $\rightarrow$ the in-sample areas are sparsely distributed spatially, leading to difficulty in constructing a spatial model at the PCS level.

Income: Distribution of areas with survey samples



Income: a possible solution

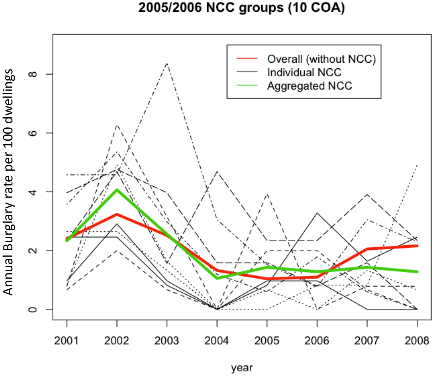
- Spatial models can be constructed at a higher geographical level, e.g., Local Authority District level (defined by the black boundary).
- All PCS's within a district assumed to share same random effect.

Outline

NCC: Background

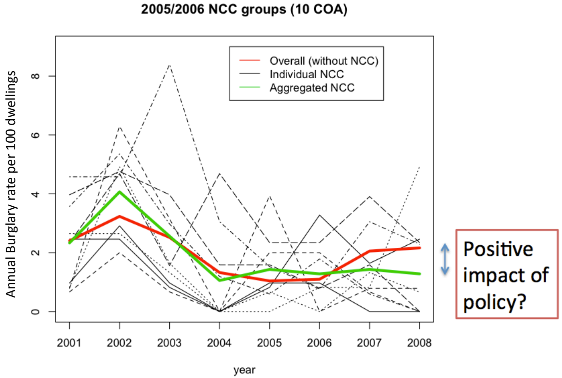
- The “No Cold Calling” (NCC) project aims to reduce the impacts of distraction burglary and rogue trading in terms of (i) the number of incidents and (ii) the public’s fear of crime.
- The scheme was initiated by the Cambridgeshire and Peterborough Distraction Burglary and Rogue Trader Task Force in 2005.
- This project was first implemented in selected areas (much smaller than a typical Census Output Area) within Peterborough.
- An evaluation of the NCC scheme by the Cambridgeshire Police concluded that residents in the NCC areas generally expressed increased confidence in dealing with unknown visitors.
- **Our aim**: to evaluate whether the NCC scheme had a measurable impact on burglary rate in the targeted areas.

NCC: trend patterns from raw data



Positive impact of policy?

NCC: trend patterns from raw data



- Before the NCC scheme (2001-2004), the overall burglary rates were similar between the NCC group and other areas in Peterborough (Rate ratio=1.06 with p-value=0.56).
- After the NCC scheme (2005-2008), the NCC group showed a reduction in the overall burglary rate (RR=0.85 and p-value=0.19).
- Raw data too sparse to draw firm conclusions about impact of scheme.

NCC: Strategy for evaluation

1. Comparing burglary rates before and after the implementation of the NCC scheme;
   - Difference between the two time periods is indicative of the influence of the policy.
2. Comparison is done after adjusting for systematic changes in burglary rates in other (non NCC) areas;
   - $\rightarrow$ The use of control areas helps to differentiate how much of the change may be due to the policy impact and how much of the change may be due to other external factors.
3. Deal with sparsity of data (i.e., small number of burglary events) by temporal smoothing and hierarchical structure across areas.
   - $\rightarrow$ separate signal from noise

NCC: Constructing the control group

- To form the control group, areas are selected on the basis of having similar local characteristics (e.g., burglary rates or deprivation scores) to those in the NCC-targeted group.

ID	Matching criterion for control group	No. of LSOAs
1	All LSOAs in Peterborough	88
2	±10% burglary rate of the NCC group in 2005	9
3	±20% burglary rate of the NCC group in 2005	20
4	±30% burglary rate of the NCC group in both 2004 and 2005	8
5	LSOAs containing the NCC-targeted COAs (but excluding the NCC-targeted COAs)	9 (one LSOA is outside Peterborough)
6	LSOAs that had "similar" multiple deprivation scores (MDS) to those for the NCC LSOAs in 2004	46

NCC: Model for the control areas

The following model is fitted to data of the control group to construct the control trend pattern.

NCC: Model for the control areas

The following model is fitted to data of the control group to construct the control trend pattern.

$$y_{it} \sim \text{Poisson}(n_i \cdot \theta_{it})$$

NCC: Model for the control areas

The following model is fitted to data of the control group to construct the control trend pattern.

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(n_i \cdot \theta_{it}) \\
\log(\theta_{it}) &= \alpha + u_i + \gamma_t + \epsilon_{it}
\end{aligned}
$$

NCC: Model for the control areas

The following model is fitted to data of the control group to construct the control trend pattern.

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(n_i \cdot \theta_{it}) \\
\log(\theta_{it}) &= \alpha + u_i + \gamma_t + \epsilon_{it} \\
\gamma_{1:T} &\sim \text{RW}_1(\mathbf{W}, \sigma_\gamma^2) \text{ (time effect)}
\end{aligned}
$$

NCC: Model for the control areas

The following model is fitted to data of the control group to construct the control trend pattern.

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(n_i \cdot \theta_{it}) \\
\log(\theta_{it}) &= \alpha + u_i + \gamma_t + \epsilon_{it} \\
\gamma_{1:T} &\sim \text{RW}_1(\mathbf{W}, \sigma_\gamma^2) \text{ (time effect)} \\
u_i &\sim \text{N}(0, \sigma_u^2) \quad \text{(place effect)}
\end{aligned}
$$

NCC: Model for the control areas

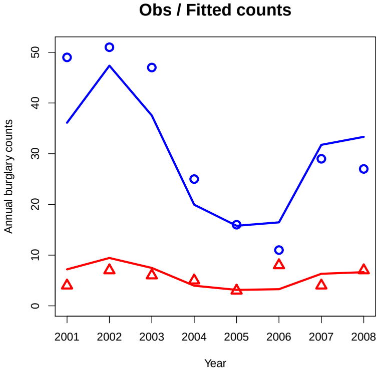
The following model is fitted to data of the control group to construct the control trend pattern.

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(n_i \cdot \theta_{it}) \\
\log(\theta_{it}) &= \alpha + u_i + \gamma_t + \epsilon_{it} \\
\gamma_{1:T} &\sim \text{RW}_1(\mathbf{W}, \sigma^2_\gamma) \text{ (time effect)} \\
u_i &\sim \text{N}(0, \sigma^2_u) \quad \text{(place effect)} \\
\epsilon_{it} &\sim \text{N}(0, \sigma^2_\epsilon) \text{ (overdispersion)}
\end{aligned}
$$

NCC: Model for the control areas

The following model is fitted to data of the control group to construct the control trend pattern.

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(n_i \cdot \theta_{it}) \\
\log(\theta_{it}) &= \alpha + u_i + \gamma_t + \epsilon_{it} \\
\gamma_{1:T} &\sim \text{RW}_1(\mathbf{W}, \sigma^2_\gamma) \text{ (time effect)} \\
u_i &\sim \text{N}(0, \sigma^2_u) \quad \text{(place effect)} \\
\epsilon_{it} &\sim \text{N}(0, \sigma^2_\epsilon) \text{ (overdispersion)}
\end{aligned}
$$

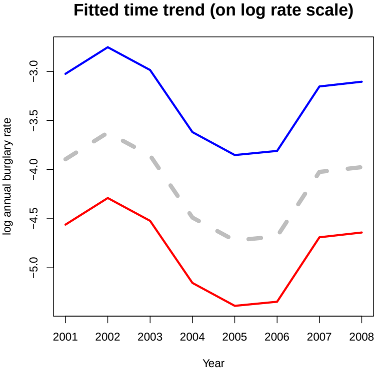


Vague priors are assigned to α and other variance parameters.

NCC: Model for the control areas

The following model is fitted to data of the control group to construct the control trend pattern.

$$
\begin{aligned}
y_{it} &\sim \text{Poisson}(n_i \cdot \theta_{it}) \\
\log(\theta_{it}) &= \alpha + u_i + \gamma_t + \epsilon_{it} \\
\gamma_{1:T} &\sim \text{RW}_1(\mathbf{W}, \sigma_\gamma^2) \text{ (time effect)} \\
u_i &\sim \text{N}(0, \sigma_u^2) \quad \text{(place effect)} \\
\epsilon_{it} &\sim \text{N}(0, \sigma_\epsilon^2) \text{ (overdispersion)}
\end{aligned}
$$

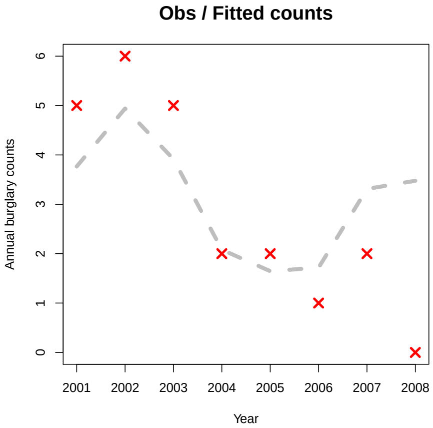


Vague priors are assigned to α and other variance parameters.

NCC: Model for the NCC areas

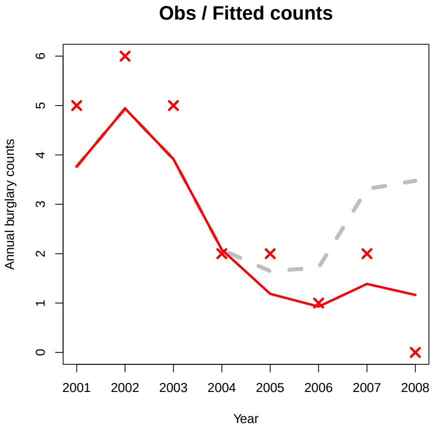
$$
\begin{aligned}
y^*_{it} &\sim \text{Poisson}(n^*_i \cdot \theta^*_{it}) \\
\log(\theta^*_{it}) &= \alpha^* + u^*_i + \gamma_t + \epsilon^*_{it}
\end{aligned}
$$

$$
\begin{aligned}
\gamma_{1:T} &\sim \text{RW}_1(\mathbf{W}, \sigma^2_\gamma) \\
u^*_i &\sim \text{N}(0, \sigma^2_u) \\
\epsilon^*_{it} &\sim \text{N}(0, \sigma^2_\epsilon)
\end{aligned}
$$



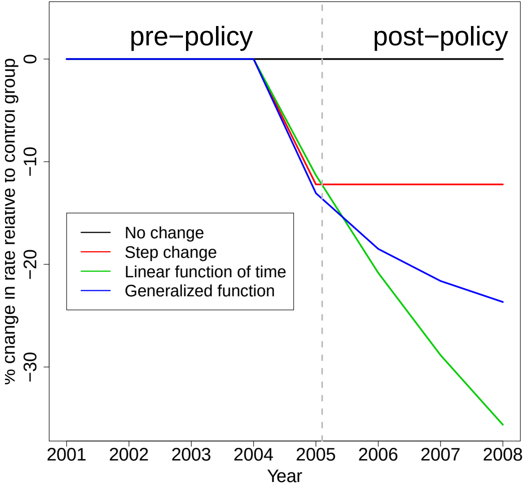
NCC: Model for the NCC areas

$$
\begin{aligned}
y^*_{it} &\sim \text{Poisson}(n^*_i \cdot \theta^*_{it}) \\
\log(\theta^*_{it}) &= \alpha^* + u^*_i + \gamma_t + \epsilon^*_{it} \\
&\quad + I(t - t_{0,i}) \cdot f(t, b_i) \\
\gamma_{1:T} &\sim \text{RW}_1(\mathbf{W}, \sigma^2_\gamma) \\
u^*_i &\sim \text{N}(0, \sigma^2_u) \\
\epsilon^*_{it} &\sim \text{N}(0, \sigma^2_\epsilon) \\
f(t, b_i) &= b_i \cdot (t - t_{0,i} + 1) \\
b_i &\sim \text{N}(\mu_b, \sigma^2_b)
\end{aligned}
$$



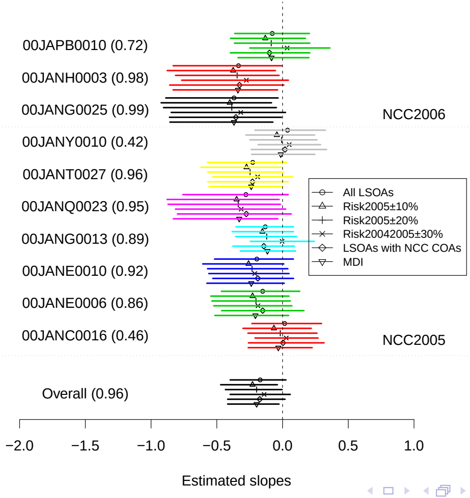
NCC: Impact functions

- We consider various functional forms for the impact function.



NCC: Local and global impacts

This figure shows the estimated slope (quantification of NCC impact) for each NCC area (b_i) and their mean (μ_b) with 95% credible intervals.



NCC: Interpretation

- NCC scheme led to an overall success
  - $\rightarrow$ an overall 16% (95% CI: -2%, 34%) reduction per year in burglary rate was estimated.
- This suggests a positive impact of the NCC policy which had the effect of "stabilizing" the burglary rate in the targeted areas while overall burglary rates were going up.
- This resulting impact function describes a gradual and persistent change and the trend of the NCC group moved further away from the control group over time.
- There exists different impacts between targeted-COAs, perhaps due to local difference in implementing the scheme.
- Estimates of burglary rates in each area borrow strength from data in other areas and time points using the hierarchical structure and hence both the local and the global impacts can be better estimated.

Outline

Summary

- In this talk, we have demonstrated some practical uses of Bayesian hierarchical models in dealing with complex data structures in small area data (spatial and temporal).
- BHM offers a natural framework for combining information which helps to strengthen inference/estimations.
- Hierarchical models with spatially/temporally structured random effects can be more reliably estimated under the Bayesian framework.
- As all variables are treated as random variables, probability statements can readily be obtained in the Bayesian framework.
- Estimation of BHM requires computationally intensive simulation methods (MCMC)
  - Implemented in free WinBUGS and GeoBUGS software: www.mrc-bsu.cam.ac.uk/bugs
  - Free software INLA (Rue et al, 2008) implements fast approximation: www.r-inla.org

Acknowledgement

- This work is funded by the ESRC National Center for Research Methods through the BIAS II project.
- We are grateful to the Swedish Statistical Bureau for providing the income data.
- The FRS project is in collaboration with the Office for National Statistics (ONS).
- We are grateful to the Cambridgeshire Constabulary and University of Cambridge (with collaboration with Prof. Robert Haining) for providing the crime data.

Reference

- V. Gomez Rubio, N. Best, S. Richardson, G. Li and P. Clarke. Bayesian Statistics Small Area Estimation.
- G. Li, R. Haining, S. Richardson and N. Best, Evaluating Neighbourhood Policing using Bayesian Hierarchical Models: No Cold Calling in Peterborough, England (in preparation)

- More information on http://www.bias-project.org.uk/