Bayesian Modelling Strategies for Spatially Varying Regression Coefficients: a Multivariate Perspective for Multiple Outcomes.

P. Congdon, Dept of Geography, Queen Mary University of London, Mile End Rd, London E1 4NS, p.congdon@qmul.ac.uk

Abstract

This paper considers modelling spatially varying regression effects for multivariate mortality count outcomes. Alternative approaches to spatial regression heterogeneity are considered: the multivariate normal conditional autoregressive (MCAR) model is contrasted with a flexible set of priors based on the multiple membership approach. These include spatial factor priors and a nonparametric approach based on the Dirichlet Process. A case study considers varying regression effects for a bivariate suicide outcome, namely male and female suicides in 354 English local authorities with social deprivation, social fragmentation and rurality as predictors.

Key Words: Spatially varying regression effects. Multiple Members Model. Conditional autoregressive priors. Multivariate Response. Suicide

1 Introduction

Bayesian disease mapping methods have focussed especially on how far health variations related to postulated unobserved risk factors may be represented by a spatially correlated error, namely a varying intercept (e.g. Besag et al, 1991), while the effects of known regressors are assumed spatially constant. In fact spatial variation in regression effects is likely (Anselin, 2001), and ignoring spatial variation in impacts of risk factors may lead to mis-statement of the average effect of covariates and neglects sub-population segmentation in the impact of risk factors.

Often multiple outcomes and multiple risk factors are involved in spatial epidemiology, and the impact of risk factors on different outcomes is correlated. The case study in this paper involves male and female suicides in English areas over 1989-93 and considers the varying impacts of social deprivation, fragmentation and level of rurality (see section 2 for a fuller discussion). These predictors are unlikely to have a spatially homogenous impact throughout England. Moreover a bivariate perspective on their spatial variability is suggested by the geographic context in which the postulated influences operate. Thus social fragmentation is greatest in highly urbanised settings with high population mobility; both male and female suicides tend to be elevated in such settings, and so one may anticipate commonality in the impacts of fragmentation on male and female suicide.

This paper accordingly seeks to generalize existing spatial univariate analysis of varying predictor effects to multiple outcomes with multiple predictors. Heterogeneity in regression effects over space is modelled by four approaches: a multivariate conditional autoregressive MCAR scheme, and three approaches (multivariate parametric, spatial factor and nonparametric with Dirichlet prior) based on the multiple membership prior characterized by underlying latent error variables.

No previous research has allowed non-parametric modelling of spatially varying regression effects in area-referenced disease mapping applications, though Gelfand et al (2004) consider a Dirichlet process approach for point-referenced data. Further, while spatial factor models have been proposed for varying spatial intercepts in multiple outcomes (Wang and Wall, 2003; Congdon, 2002), or for pooling information over spatially correlated social indicators (Hogan and Tchernis, 2004), a spatial factor approach is applied here to spatially varying regression coefficients. This is a novel approach and needed when there are similar patterns of variation in health or mortality outcomes in relation to potential social and environmental risks.

The following sections review the case study, the multivariate CAR applied to multiple outcomes, and the three variations on the multiple membership prior mentioned above. A worked analysis of the case study follows and then a final discussion reviews prospects for extensions of the approaches considered.

2 Suicide in 354 English Areas

As a case study, the present paper considers varying regression effects for a bivariate suicide outcome, namely male and female suicides in 354 English local authorities during a five-year period 1989-93. Previous work on geographical suicide variations has highlighted the impact of factors that are associated with elevated psychiatric morbidity in general, especially social deprivation (Gunnell et al, 1995). Evidence from studies such as Dorling & Gunnell (2003) is that while high area deprivation is generally associated with high suicide, certain areas have low suicide rates in view of their social structure.

Recent ecological analyses of suicide data have also shown excess risk associated with social fragmentation and social isolation. In terms of community characteristics fragmentation is associated with above average numbers of one person and other non-family households, with high population turnover, and extensive private renting in 'bedsitters'. An index summarising such factors is used by Whitley et al (1999) and Congdon (1996) to analyse suicide variations. Social fragmentation may occur in affluent areas (e.g. central London) as well as deprived areas, and deprivation and fragmentation are not necessarily highly correlated. Fragmentation scores tend to be high in inner city areas and in coastal resorts with transient workforces.

Finally, while suicide rates tend to be higher in cities, recent suicide trends also highlight the impact of isolation in certain more rural settings (Middleton et al, 2003). However the impact of rurality is complex (see Hawton et al, 1999), and a spatially constant effect of a rurality index on suicide variations is unlikely.

Therefore the analysis involves three predictors which are scores summarising the above social structural features, specifically sums of standardised z scores on constituent 1991 Census indicators: a social fragmentation score, a deprivation score and a rurality score, each with a potentially spatially varying effect. The totals of the z scores are themselves standardised so that the three predictors have zero mean, and variance unity.

Social fragmentation combines people not married, population turnover, one person households and 'bedsitter' (private rented) accommodation. Deprivation combines unemployment, social groups IV and V (semi & unskilled workers), no car households and social renting. Rurality is based on two variables: agricultural workers as a percent of the workforce (z score positively signed) and population density (z score negatively signed).

3 Univariate Models

We now consider modelling approaches for mortality and health outcomes such as the suicide data just considered. Mixed intercept models for univariate spatial health outcomes are well established. Besag et al (1991) proposed a mixed model involving unstructured effects u_i and conditionally autoregressive spatial effects s_i, that together define a random intercept for area i (i=1,..,n). For y_i a univariate mortality count, with log link and mean $E_i\mu_i$ (E_i are expected deaths), the mixed model is

$$\log(\mu_i)=\gamma_0 + X_i'\gamma + u_i+s_i \qquad (1)$$

where γ is $P\times 1$, and X_i is a $P\times 1$ vector of explanatory variables. This type of model is often used for smoothing relative mortality risks when mortality counts are small and when a spatial correlation in relative risks is expected on substantive grounds.

Let c_{ij} denote spatial interaction terms, with $c_{ii}=0$. These define the pairs of neighbouring regions and the strength of adjacency, i.e. the weights with which the neighbours act in the conditional autoregression. They may be based on distances between areas or simply contiguity, and may contain unknown parameters (e.g. for distance decay).

Under contiguity, with c_{ij}=1 if areas i and j are adjacent, and 0 otherwise, the conditionally autoregressive (CAR) prior of Besag et al (1991) has the form

$$s_i|s_{[i]} \sim N(S_i,\varphi/N_i) \qquad (2)$$

where $s_{[i]}$ is the set $\{s_1,\ldots s_{i-1},s_{i+1},..,s_n\}$ excluding s_i,

$$S_i = \sum_{i\neq j}^{n} c_{ij}s_j/\sum_{i\neq j}^{n} c_{ij}$$

is the average of s_j for the j=1,..N_i neighbours of area i, and φ is a variance parameter. A fixed effect intercept may be included in (1) only if the s_i are centred to have mean zero. Correlation between u_i and s_i is not possible to incorporate under the CAR prior for s_i, and can only be assessed by the empirical correlation between posterior estimates $u_i|y$ and $s_i|y$.

The preceding model is one for spatially correlated residuals and sometimes proposed as representing the impact of unmeasured and spatially correlated risk factors. An alternative source of spatial patterning in residuals is through spatially varying predictor effects, and spatial correlation in the effects of known predictors (see Fotheringham et al, 2003, for real data examples). Spatially varying coefficient models may be relevant when the totality of areas includes different contextual settings (e.g. urban and rural areas) where different predictor mechanisms underlie variation in relative risks. For example, the causation of high suicide rural areas may not be the same as for high suicide inner city areas.

One possible univariate model with spatially varying regression coefficients is

$$\log(\mu_i)= \gamma_0 + X_i'\gamma_i \qquad (3)$$

where $\gamma_i=(\gamma_{i1},\ldots\gamma_{iP})'$ vary by area according to a multivariate version of the CAR prior. Specifically, the MCAR prior of Gamerman et al (2003) generalises the univariate pairwise difference joint prior so that

$$P(\gamma_1,\ldots,\gamma_n|\Phi_\gamma) \propto |\Phi_\gamma|^{-0.5n} \exp[-0.5\sum_{i\neq j}^{n} c_{ij}(\gamma_i - \gamma_j)'\Phi_\gamma^{-1}(\gamma_i-\gamma_j)]$$

where Φ_γ is a dispersion matrix of order P. The conditional autoregressive form of this prior under contiguity adjacency is

$$\gamma_i|\gamma_{[i]} \sim N_P(\overline{\Gamma}_i, \Phi_\gamma/N_i) \qquad (4)$$

where $\overline{\Gamma}_i =[\overline{\gamma}_{i1,\ldots\ldots}, \overline{\gamma}_{ip}]$ and $\overline{\gamma}_{ip}$ is the average of γ_{jp} over the $j=1,..,N_i$ neighbours of area i for predictors p=1,..,P.

Proper priors need to be specified for Φ_γ. A Wishart prior with scale matrix C_γ and ν_γ degrees of freedom for Φ_γ^{-1} is one option (Gamerman et al, 2003, p 515), though priors for covariance matrices that facilitate incorporation of prior knowledge regarding correlated effects have been proposed (Barnard et al, 2000). One might provide a prior estimate of the covariance matrix C_γ in the form $C_\gamma=D_\gamma R_\gamma D_\gamma$ where $D=\text{diag}(\sigma_{\gamma_1},\ldots, \sigma_{\gamma_P})$ is a diagonal matrix containing prior estimates of the standard deviations σ_{γ_p} of γ_{ip}, and $R_\gamma=[r_{km}]$ is a prior estimate of the matrix of correlations between γ_{ik} and γ_{im}. Another possibility would be to simulate from different candidate priors on the basis of known predictors and expected deaths but not referring to the actual responses; this is discussed and illustrated in section 5 (on parametric multiple membership priors). Defaults for C_γ are often used such as the identity matrix, in which case ν_γ typically takes the default value $\nu_\gamma=P$ (e.g. Chib and Winkelmann, 2001, p 431). A more informative estimate for C_γ might justify a larger value for ν_γ though in large datasets one would expect the data to outweigh the prior unless it is fairly informative.

4 MCAR models generalized to Multiple Outcomes

The above models relate to a univariate outcome, but multivariate outcomes are frequent in spatial epidemiology and there may be benefits in pooling information over outcomes which are substantively interrelated, e.g. the same mortality cause but disaggregated to different genders or ethnic groups. Typically one expects both correlation in effects of particular predictors across outcomes but also distinctive features in each outcome's patterning. For instance, on average fragmentation may have a stronger effect on female than male suicide, and deprivation a stronger effect on male suicide.

Suppose multivariate observations of dimension M are observed for each area. In particular, let the responses for n areas consist of multivariate counts

$$\{Y_1,\ldots,Y_n\} = \{y_{11},y_{12},..y_{1M};y_{21},y_{22},..y_{2M};\ldots y_{n1},y_{n2},\ldots y_{nM}\}$$

with $Y_i=(y_{i1},\ldots y_{iM})$, which might for instance be data on deaths by gender or by cause group. Assuming, at least initially, Poisson variation, a varying intercept model parallel to (1) for the Poisson means μ_{im} of y_{im} is

$$\log(\mu_{im})=\beta_{0m}+X_i'\beta_m+u_{im}+s_{im} \qquad (5)$$

where the s_{im} are spatially dependent errors, which may follow an MCAR prior, and the covariate effects $\beta_m=(\beta_{1m},\ldots\beta_{Pm})$ are specific to outcome m but spatially constant. However, suppose spatially varying predictor effects are also allowed with multivariate outcomes as in

$$\log(\mu_{im})=\beta_{0m}+X_i'\beta_{im}+u_{im}+s_{im} \qquad (6)$$

where $\beta_{im}=(\beta_{i1m},\ldots\beta_{iPm})$. So the covariate effects for area i, namely

$$B_i=(\beta_{i11},\ldots\beta_{iP1},\beta_{i12},\ldots\beta_{iP2},\ldots.\beta_{i1M},..\beta_{iPM})$$

$= (\beta_{i1},\ldots\beta_{iM})$

are outcome and predictor specific. A reduced model

$$\log(\mu_{im}) = \beta_{0m} + X_i'\beta_{im} \qquad (7)$$

may be used if spatially varying regression effects remove spatial correlation in the intercept errors. Under (7) there may be

a) correlations $r(\beta_{pm},\beta_{qm})$ between the effects β_{ipm} and β_{iqm} of predictors {p,q} on the m^{th} response,

b) correlations $r(\beta_{pm_1},\beta_{pm_2})$ between the effects β_{ipm_1} and β_{ipm_2} of the p^{th} predictor on different responses variables, and

c) cross-correlations between different predictors and different variates $r(\beta_{pm_1}, \beta_{qm_2})$ for $p\neq q, m_1\neq m_2$.

The MCAR prior now has the form

$$P(B_1,\ldots, B_n|\Phi_\beta) \propto |\Phi_\beta|^{-0.5n}$$

$$\exp[-0.5\sum_{i\neq j}^{n} c_{ij}(B_i-B_j)'\Phi_\beta^{-1}(B_i-B_j)] \quad (8)$$

where Φ_β is of order MP. As for Φ_γ in the univariate case, one may specify a Wishart prior for Φ_β^{-1} with scale matrix C_β and ν_β degrees of freedom. A prior estimate of the covariance matrix C_β of the form $C_\beta=D_\beta R_\beta D_\beta$ could be provided (including information on the three forms of correlation mentioned above), and this could justify a larger degrees of freedom than the default ν_β=MP. The conditional autoregressive form for multioutcome-multivariate predictor effects under binary adjacency is

$$B_i | B_{[i]} \sim N_{MP}(\bar{B}_i, \Phi_\beta / N_i)$$

where

$$\bar{B}_i = [\bar{\beta}_{i11,\ldots,}\bar{\beta}_{iP1},\bar{\beta}_{i12,\ldots,}\bar{\beta}_{iP2,}\ldots,\bar{\beta}_{i1M,\ldots.}\bar{\beta}_{iPM}],$$

and $\bar{\beta}_{ipm}$ is formed as the average of β_{jpm} over the j=1,..,N_i neighbours of area i. The overall average effect $\bar{\beta}_{pm}$ of predictor p on response m is the average over areas of the effects $\bar{\beta}_{ipm}$.

This and subsequent models involve sets of random effects with the model dimension not known precisely. Model fit and dimension may be assessed by the method of Spiegelhalter et al (2002) based on the DIC (analogous to the established AIC but suitable for random effects models). This method provides a measure of model complexity d_e. Useful as a measure of fit is the log pseudo marginal likelihood (PsML) of Gelfand (1996); the individual components of this are Monte Carlo estimates of the log CPOs and useful in indicating poorly fitted cases. An model fit criterion focussed on model predictions is provided by the squared error loss predictive criterion of Gelfand and Ghosh (1998). Suppose replicate data Z_{im} are sampled from the posterior predictive density $P(Z_m|Y)$, where $Z_m=(Z_{1m},Z_{2m},\ldots\ Z_{nm})$. Then with $\varsigma_{im}=Var(Z_{im})$ and $\upsilon_{im}=E(Z_{im})$, the squared error loss criterion for a particular value of c is

$$PL_c=\sum_{i=1}^{N}\sum_{m=1}^{M}\varsigma_{im}+c/(c+1)\sum_{i=1}^{N}\sum_{m=1}^{M}(y_{im}-\upsilon_{im})^2 \qquad (9)$$

with large c values downplaying precision of predictions. These techniques are often used (in an informal Bayes sense) to prefer one model against another.

Cross-validatory predictive checks for individual areas in this and subsequent multivariate models are based on the mixed predictive approach of Marshall and Spiegelhalter (2003). Let $Z^*_m=(Z^*_{1m},Z^*_{2m},\ldots,Z^*_{nm})$ be posterior replications under this approach. Then the predictive check is whether the probability

$$Pr(Z^*_{im}> y_{im})+0.5Pr(Z^*_{im}=y_{im})$$

is under 0.025 or exceeds 0.975. This procedure assists in detecting possible outlying cases (cases not well fitted by a model) and is less conservative than other Bayesian predictive testing approaches. Additionally, the Moran I statistic is calculated from model residuals $y_{im}-\hat{y}_{im}$ to assess whether spatial correlation is removed from the regression residuals.

5 Methodology: Multiple Member Priors for Multivariate Outcomes

Another possible approach proposed here for multi-outcome multivariate spatial regression effects extends the spatial intercepts method of Leyland et al (2000), Browne et al (2001) and Langford et al (1999) to obtain spatially varying regression coefficients. This approach, also known as the multiple members approach, and has links to moving average error models (e.g. Best et al, 2002). It was proposed in a frequentist framework and for varying intercepts, but is here framed as a Bayesian prior applicable to varying predictor effects and denoted the multiple members prior (MMP).

Consider a multivariate count response (y_{i1},y_{i2},… y_{iM}) with means ($E_{i1}\mu_{i1}$, $E_{i2}\mu_{i2}$,…,$E_{iM}\mu_{iM}$), and the model in (6) with two varying intercepts as well as varying effects on P predictors. In the most general formulation, a multiple members prior that defines all random effects (both random regression coefficients and random intercepts) in the conditional mean for y_{im} is possible.

Thus consider an underlying vector of random effects V_i = (v_{i1},v_{i2},…v_{iM}), where v_{im} = (v_{i1m},v_{i2m},…$v_{i,P+2,m}$) is of dimension P+2 and defines the values of [β_{im},s_{im},u_{im}], where β_{im}=(β_{i1m},β_{i2m},…β_{iPm}) as above. So V_i corresponds to

$\{[\beta_{i1},s_{i1},u_{i1}],[\beta_{i2},s_{i2},u_{i2}],\ldots[\beta_{iM},s_{iM},u_{iM}]\}$

The predictor effects β_{ipm} in the conditional mean μ_{im} for y_{im} are obtained as a spatially weighted average $\beta_{ipm} = \sum_{i=1}^{n} w_{ij} v_{jpm}$, where $w_{ij}=c_{ij}/\sum_{i\neq j}^{n} c_{ij}$, and the spatial effect is obtained as $s_{im} = \sum_{i=1}^{n} w_{ij}\, v_{i,P+1,m}$. Under contiguity these reduce to $\sum_{j\in L_i} v_{jpm} /N_i$ and $\sum_{j\in L_i} v_{i,P+1,m} /N_i$ where L_i denotes the set of neighbours of area i (N_i in total). Unstructured effects in μ_{im} are obtained without transformation, so that $u_{im}=v_{i,P+2,m}$.

There is no reduction in the (nominal) dimension of the random effects under this option, though in random effects models such as this, the effective dimensionality is typically much less than the nominal dimensionality; the method of Spiegelhalter et al (2002) is one method to estimate the effective dimensionality. However, as following sections demonstrate, there is a considerable flexibility gained under the MMP approach from the fact that the underlying errors are unstructured. This facilitates nonparametric or discrete mixture priors for v_{im} or common factor methods of lower dimension than M(P+2). Hence this form of prior could be used as an alternative to the MCAR (which assumes symmetric normal predictor effects) when there is some doubt about the shape of predictor effects.

The multiple members approach also allows the v terms underlying spatially unstructured intercepts or predictor effects to be correlated with the corresponding spatially structured effects (this is not possible under CAR approaches). In particular, the MMP prior facilitates models including unstructured but varying predictor effects $\delta_{im}=(\delta_{11m},\delta_{i2m},\ldots\delta_{iPm})$, with V_i then of dimension 2P+2, $v_{im}=[\beta_{im},\delta_{im},s_{im},u_{im}]$, and conditional mean for y_{im} as

$$\log(\mu_{im})=\beta_{0m}+X_i'(\beta_{im}+\delta_{im})+u_{im}+s_{im}\ .$$

Consider, however, the model $\log(\mu_{im})=\beta_{0m}+X_i'\beta_{im}+u_{im}+s_{im}$. Under a parametric MMP prior, $V_i = (v_{i1},v_{i2},\ldots v_{iM})$ might be taken as multivariate normal with prior mean m_v of the form $[(\kappa_1,0,0), (\kappa_2,0,0),.. (\kappa_M,0,0)]$, where $\kappa_m=\{\kappa_{1m},\ldots\kappa_{Pm}\}$ is a vector of length P containing mean effects of predictor p ($\in$1,..P) on outcome m. The κ_{pm} are assigned fixed effects priors. The covariance matrix Φ_v is of order MP+2M and a Wishart prior may be assumed, with scale matrix provided by a prior covariance estimate C_v. If predictors are on the same scale (e.g. standardised) it seems reasonable to assume equal prior standard deviations for the effects $(v_{i1m},v_{i2m},\ldots v_{i,P+2,m})$ underlying the regression coefficients and intercepts. So C_v has the form k_1I where I is an MP+2M identity matrix and k_1 can be varied, with resulting variation in conditional means μ_{im}, or expected relative risks μ_{im}/E_{im}, tested against plausible variations in the mortality outcome. In spatial mortality applications there is often accumulated substantive evidence about plausible variations in relative risks (RRs) between areas. For example, in the set of English local authorities (with typically around 200 thousand population) variations in suicide relative risk (about a mean of 1) of more than twenty (e.g. from RR=0.2 to RR=4) are a priori unlikely. In fact the variations in crude relative risk (suicide standard mortality ratios in the 354 English areas over 1989-93) are much less than this: the 0.025 and 0.975 percentiles of male & female SMRs are (0.66,1.5) and (0.42, 2.16) respectively. In the case of prior mean predictor effects κ_{pm} (the first P elements in the prior mean for v_{im}), simulation relating to C_v is combined with simulation from a normal prior $\kappa_{pm} \sim N(0,k_2)$.

As an example, combinations of (k_1,k_2) were applied to the case study data (as section 2 mentions, the three predictors are standardised), and the 0.025 and 0.975 points in the model RRs, namely μ_{im}/E_{im} (m=1,2), are monitored, these points being denoted $\{\eta_{Lm},\eta_{Hm}\}$. This procedure involves simulation from the priors only, not the actual y_{im}, and is found to yield approximately symmetric densities for η_{Lm} but positively skewed densities for η_{Hm}. For k_1=0.001 and k_2=2.5 the means (medians) for η_{Lm} are 0.36 (0.34) for males and 0.35 (0.33) for females, while the means (medians) for η_{Hm} are 12.8 (2.9) for males and 5.2 (2.9) for females. Such settings more than encompass the observed range in relative risks and might be used to set mildly informative priors for κ_{pm} and Φ_v.

In the suicide mortality case study, M=2 and P=3, so V_i in a model with spatially structured predictor effects and varying intercepts (u_i and s_i) has dimension M(P+2)=10. The underlying effects are linked to the spatially varying regression coefficients and intercepts as follows:

$$\beta_{i11} = \sum_{j=1}^{n} w_{ij} v_{j11}, \qquad (10)$$

$$\beta_{i21} = \sum_{j=1}^{n} w_{ij} v_{j21},$$

$$\beta_{i31} = \sum_{j=1}^{n} w_{ij} v_{j31},$$

$$s_{i1} = \sum_{j=1}^{n} w_{ij} v_{j41},$$

$$u_{i1} = v_{i51},$$

$$\beta_{i12} = \sum_{j=1}^{n} w_{ij} v_{j12},$$

$$\beta_{i22} = \sum_{j=1}^{n} w_{ij} v_{j22,}$$

$$\beta_{i32} = \sum_{j=1}^{n} w_{ij} v_{j32},$$

$$s_{i2} = \sum_{j=1}^{n} w_{ij} v_{j42},$$

$$u_{i2} = v_{i52}$$

Then one might take $V_i = (v_{i11},v_{i21},\ldots v_{i52})$ to be parametric random effects, e.g. multivariate normal of order 10 with prior dispersion matrix Φ_v and prior mean $m_v=(\kappa_{11},\kappa_{21},\kappa_{31},0,0,\kappa_{12},\kappa_{22},\kappa_{32},0,0)$. The prior means of the effects $\{v_{i4},v_{i5,}v_{i9},v_{i10}\}$ underlying the varying spatial intercepts are zero, but the remainder are prior estimates κ_{pm} (p=1,..,P; m=1,..,M) of outcome specific predictor effects. A multivariate Student t model might be used involving gamma distributed scale factors ω_i specific to each area. This might be used to model spatial heteroscedasticity or outlier areas.

It may be possible to omit intercept variation if the response specific model deviations $y_{im}-\hat{y}_{im}$ are not spatially correlated. Then V_i reduces to dimension MP and v_{im} of dimension P defines the values of β_{im}. In the case study $V_i = (v_{i1},v_{i2},\ldots v_{i6})$ might be multivariate Normal or Student with mean $m_v=(\kappa_{11}, \kappa_{21}, \kappa_{31}, \kappa_{12}, \kappa_{22}, \kappa_{32})$.

6 Methodology: Spatial Factor Model based on Multiple Members Prior

With multiple correlated outcomes it is likely that predictors will have similar effects on the outcomes (e.g. positive effects of deprivation and fragmentation on both male and female suicide). By extension if predictor effects are allowed to vary spatially, the spatial patterning of regression effects will be expected to show similarities across outcomes. Hence a spatial factor approach to spatially varying coefficients is proposed that seeks to provide a parsimonious parameterisation when there is correlation between spatially varying predictor effects over outcomes. This involves a straightforward modification of the priors in section 5, and might be applicable when mortality outcomes are strongly correlated in their patterns of geographic prevalence and in their patterns of association with ecological risk factors.

The common factor (or possibly factors) may model correlations between effects of predictors k and/or between outcomes m. For example, predictor specific factors $\{F_{ipr}, r=1,\ldots R\}$, where $R<M$, might model the correlations $r(\beta_{pm_1},\beta_{pm_2})$ between the effects β_{ipm} of predictor p on different outcomes. Outcome specific factors G_{imu}, $u=1,..U$ ($U < P$), might model the correlations $r(\beta_{pm},\beta_{qm})$ between the effects β_{ipm} and β_{iqm} of predictors {p,q} on the m^{th} response. Factors H_{is}, $s=1,..S$, where $S < MP$, might model the cross-correlations $r(\beta_{pm_1}, \beta_{qm_2})$ between β_{ipm} over both different predictors (p,q) and different variates (m_1,m_2).

As an example suppose R=1, and F_{ip} is the underlying factor score for area i and predictor p. So there are P sets of n scores with the score vector for area i being $F_i=\{F_{i1},..F_{iP}\}$. Then the spatially varying regression coefficients β_{ipm} are obtained by applying loadings λ_{mp} to spatially filtered sums of factor scores,

$$\beta_{ipm} = \lambda_{mp} \sum_{j=1}^{n} w_{ij} F_{jp} \tag{11}$$

For the case R=1, $F_i=(F_{i1},\ldots,F_{iP})$ is of dimension P, and might be modelled as multivariate Normal with non-zero means $m_F=(\kappa_1, \kappa_2,\ldots,\kappa_P)$. If the variances/covariances of F_{ip} in the adopted multivariate density are left as free parameters, the loadings λ_{1p} may be set (e.g. to 1) for identifiability, leaving (M-1)P free loadings.

Alternatively with U=1, outcome specific factor scores G_{im} account for spatial regression coefficients as

$$\beta_{ipm} = \xi_{pm} \sum_{j=1}^{n} w_{ij} G_{jm} \tag{12}$$

where $G_i=(G_{i1},\ldots G_{iM})$ is possibly taken as MVN of dimension M.

7 Methodology: Nonparametric Regression Heterogeneity under the Multiple Members Prior

A final proposal here is for robust spatial regression models based on a nonparametric approach applied to the multiple members prior. The mixed model of Besag et al (1991) and modifications of it that seek to model discontinuities (e.g. Lawson and Clark, 2002) have considered intercept variation via spatial 'errors' and not variation in regression effects. As one option to extend robust models to spatial regression effect models, one may let the underlying V_i in section 5 be drawn from a discrete mixture with a known number of components.

Alternatively, following Ishwaran and Zarepour (2002a, 2002b), a truncated version of the infinite Dirichlet process (DP) prior of Ferguson (1973, 1974) is applied. In general terms the DPP approach takes the distribution function F underlying a parameter θ to be unknown (whereas parametric approaches assume F is known). F is centred on a baseline density G according to a concentration parameter α, so that the larger is α the closer F is to G. For a univariate parameter, the real line is partitioned into disjoint intervals $(-\infty,u_1),..,[u_{j-1},u_j),..[u_k,\infty)$ with probabilities $p_j=Pr(\theta\in[u_j,u_{j+1}))=F(u_j)-F(u_{j-1})$. Then the probabilities $p_{Gj}=G(u_j)-G(u_{j-1})$ under G define the prior for the p_j, such that

$$(p_1,\ldots.p_k) \sim Dir(\alpha p_{G1},\ldots.\alpha p_{Gk}).$$

The constructive definition of the Dirichlet prior (Sethuraman, 1994) generates the p_j using a stick breaking prior (SBP) whereby $r_j \sim Be(1,\alpha)$, $p_1=r_1$, $p_2=(1-r_1)r_2$, $p_3=(1-r_1)(1-r_2)r_3$,..,etc. The constructive definition shows that sampling θ from G and p from the SBP leads to samples from an infinite mixture of point masses (Muller & Quintana, 2004)

$$h(\theta) = \sum_{k=1}^{\infty} p_k I(\theta_k)$$

This makes the original form of the DP prior unsuitable when some smoothness in the underlying density of the parameter(s) is expected. However, one may adopt an alternative framework (e.g. West et al, 1994; Mukhopadhyay & Gelfand, 1997) involving a mixture of parametric densities f(|θ), or DP mixture (MDP) model (Green & Richardson, 2001), whereby

$$h(\theta) = \sum_{k=1}^{\infty} p_k f(|\theta_k), \qquad \theta_k \sim G$$

Furthermore MCMC sampling is simplified by truncating the number of possible clusters from which candidate values for a parameter or parameters can be drawn (Ishwaran and Zarepour, 2002a; 2002b; Mueller & Quintana, 2004). So with K clusters,

$$\sum_{k=1}^{\infty} p_k f(|\theta_k) \approx \sum_{k=1}^{K} p_k f(|\theta_k),$$

In the present spatial application involving the multiple members prior, consider the case where V_i consists only of spatially varying predictor effects (i.e. without varying intercepts included in v_{im}). Under a truncated MDP approach, V_i is selected from K possible cluster values V_k^*, following a baseline prior G_K, where the number of clusters K is less than or equal to n. The baseline prior is appropriate to modelling potentially smooth unimodal random effects (e.g. G_K is an MVN prior for V_k^*), with the closeness of F to G_K increasing as $\alpha \to \infty$ (Pretorius & van der Merwe, 2002). However, the process allows for departures from smoothness caused by clumping, multimodality, etc. So, as in Ishwaran and Zarepour (2002b, section 2)

$$y_{im}| V_i=V^*_k \sim Po(E_{im}\mu_{im})$$

$$\log(\mu_{im})=\beta_{0m}+X'_i\beta^*_{km}$$

$$V^*|F \sim F$$

$$F \sim DP(\alpha G_K)$$

where G_K consists of a multivariate normal (of dimension MP) for V*, with means $m_{V^*_k}$ of the cluster priors allowed to differ, but a common dispersion matrix Φ_{V^*} assumed. The translation from V^*_k to β^*_{km} is as described in section 4.

Equivalently (Ishwaran and Zarepour, 2002b, sections 2.1 and 3), the process can be characterised by classification or 'configuration' indicators $D_i \in (1,\ldots K)$ where the D_i identify which cluster k (i.e. which V^*_k) is selected for area i. The link between partition and allocation priors for DP and other nonparametric models is demonstrated by Green & Richardson (2001).

For example, with M=2 outcomes, and P=3 predictors, G_K might be a multivariate Normal or Student t of dimension 6. If at a particular iteration t=1,..,T in a MCMC sampling chain, the selected cluster for area j is $c_j=D^{(t)}_j$, one obtains $\beta^{(t)}_{ipm}$ as a spatially filtered average of v^{*}_{cjpm}. The mean predictor effects $\{\bar{\beta}_{11}, \bar{\beta}_{21}, \bar{\beta}_{31}, \bar{\beta}_{12}, \bar{\beta}_{22}, \bar{\beta}_{32}\}$ are obtained by averaging the $\beta^{(t)}_{ipm}$ over areas and iterations. The actual number of non-empty clusters $J^{(t)} \leq K$ will vary over iterations. The WINBUGS coding of DP models is considered by Congdon (2001) and his code is adapted as an example at the WINBUGS website.

8 Case Study: Model Sequence and Model Assessment

The observations y_{im} on male and female deaths (m=1,2) are assumed to be Poisson with means $E_{im}\mu_{im}$, with expected suicides E_{im} obtained by using England age specific rates for 1991. Implementation involves the WINBUGS package (Spiegelhalter et al, 2003). The baseline model (model A) involves univariate Poisson regression models without varying intercepts, namely

$$\log(\mu_{im})=\beta_{0m}+X_i'\beta_m \qquad m=1,..,M$$

where X_i is a P×1 predictor for area i, with M=2 and P=3. The eight regression parameters in this model are assigned N(0,1000) priors. The first two heterogeneous regression models (B and C) are fully parametric and multivariate (pooling random effects over the two outcomes). They follow the MCAR and MMP specifications in sections 4 and 5 respectively, and make the assumption (which is subject to assessment) that spatial intercepts are not needed when varying regression effects are allowed. This involves applying spatial correlation tests to residuals.

In model B, the precision matrix Φ_β^{-1} is assigned a Wishart prior with identity scale matrix and six degrees of freedom; this is a common default prior in multivariate count models (Chib & Winkelmann, 2001). The same prior applies to Φ_v^{-1} in model C. Unlike in model A, in models B and C (and subsequent models including random effects), the number of parameters is unknown, but the complexity measure d_e is also interpretable as a measure of effective model dimension (Spiegelhalter et al, 2002).

Models D1 and D2 are spatial factor models. Model D1 assumes predictor specific factors F_{ip}, p=1,..,P; so R=1 as in (11) above. These factor scores are assigned a trivariate Normal prior $N_3(m_F,\Phi_F)$ with unknown dispersion matrices, so the loadings for females λ_{2p} are assigned N(1,1) priors but those for males are set to 1. It is possible to assume a diffuse prior for the loadings but we follow the mildly informative N(1,1) prior suggested by Johnson and Albert (1999, p 197). Given that λ_{1p}=1 we expect on substantive grounds the female loading to be centred at unity. The fixed effects κ_p (p=1,..,P) in m_F are assigned N(0,1000) priors. Model D2 assumes outcome specific factors G_{im}, with U=1 as in (12), with (G_{i1},G_{i2}) taken as bivariate Normal with unknown dispersion matrix. Hence the deprivation loadings ξ_{1m} are set to 1 for identifiability.

Model E is a nonparametric spatial regression model with the maximum possible number of clusters K set at 20 and κ set at 0.5. A multivariate Normal baseline prior G_K of dimension 6 for the clustered regression effects is assumed, with the prior on Φ_v^{-1} as in model C. Preliminary analysis over a small range of alternative values (κ=0.5, 1 and 5) suggested a relatively small value of κ led to a lower DIC, whereas κ=5 led to increased effective model dimension d_e. Additionally taking κ as a free parameter, namely $\kappa \sim Ga(10,10/a)$, with a=1 being a prior estimate of the mean of κ (cf. West and Turner, 1994), led to a posterior mean for α of just over 0.5.

Finally, model F considers the univariate equivalent of whichever of models B to D appears most appropriate to the observations. Thus random effects under model F allow for correlations between predictor effects within each outcome but there is no pooling strength over outcomes. This is to assess the benefit of a multivariate approach to the case study data as argued in the Introduction. Please note that code for these models is available from www.geog.qmul.ac.uk/staff/congdon.html.

9 Case Study: Results

Estimation uses two chains with dispersed initial values run for 2500 iterations. Convergence was assessed by Gelman-Rubin criteria (Gelman et al, 2003) and in all models obtained by iteration 1000, with inferences based on iterations 1001-2500.

We find from model B that average predictor effects $\bar{\beta}_{pm}$ are all significantly positive for males, but that only social fragmentation has a clear positive effect on female suicide (Table 1). The same pattern applies for the Poisson regression model (model A). All predictors are in standard form; so the median fragmentation coefficient of 0.136 for females implies a relative risk of 1.31 for the areas with a fragmentation score of 1.96. In fact this predictor shows some positive skewness, with 12 areas having predictor values over 2.5.

It is also apparent that whereas model A deviations are spatially correlated, under model B the Moran I statistics for male and female suicides are not different from zero, so there is no residual spatial correlation. This confirms that varying intercepts are not needed to account for residual spatial correlation; allowing spatial heterogeneity in regression impacts is sufficient to remove spatially correlated residuals.

Mean regression effects under the fully parametric MMP model C are similar to those of model B though the fit appears to improve slightly (see Table 2). The correlation between the 354 values of β_{i11} under model B and under model C is 0.75. For the five other coefficients (β_{i21} through to β_{i32}) the correlations are 0.77, 0.78, 0.80, 0.77 and 0.88. So the spatially varying regression coefficients seem to be reasonably stable over the two models. The Moran coefficient is also satisfactory under model C. The effective parameter count is reduced under model C as compared to model B, and the predictive loss measure in (9), which also includes a complexity penalty, is lower for model C.

By contrast, the Marshall and Spiegelhalter (2003) diagnostics show more divergencies (P-values over 0.975 or under 0.025) for model C than model B, namely 26 vs. 18 in 758 observations (N=354 male suicide totals, N=354 female suicide totals). However, a Q-Q plot of the P values for model C against the 2N order statistics 1/(2N+1) of a uniform (0,1) distribution appears satisfactory; Figure 1 shows very similar plots as between models B and C.

The spatial factor models D1 and D2 have fewer effective parameters as would be expected. The predictor specific factor model D1 has a lower DIC than models B and C and a lower value of the PL criterion (for c=1). There are 35 divergent areas as per Marshall and Spiegelhalter (2003), though the Q-Q plot is in line with a uniform distribution. The outcome specific factor model D2 fits far less well, with deterioration on both model fit and model checks. Both Moran statistics have 95% intervals above zero. Table 1 only shows the regression effect summary for model D1.

Model E, the nonparametric MMP model, produces a better DIC than the other varying coefficient models, though the Moran statistic for females is possibly problematic. However, the 2.5% point is effectively zero. As for model D1, deterioration in $Dev(\bar{\theta})$ as compared to models B and C is offset by a reduced effective parameter total. This model is distinctive in reducing divergent areas as per Marshall and Spiegelhalter (2003) to only four. This is as might be expected, namely that a non-parametric model is better adapted to accommodate atypical values. A fully parametric model may tend to shrink extreme relative risks, and so mask unusual observations that do not conform to a smooth unimodal pattern. The expected prior number of clusters with $\alpha=0.5$ is approximately $\alpha\log(1+354/\alpha)=3.28$ (West and Turner, 1994), whereas the posterior mean number of clusters was 5.1.

It is apparent that while model E is perhaps preferred on the grounds of the model fit criteria, models B and C more clearly eliminate spatial correlation in the errors. So on model checking and fit criteria considered together, model C may be preferred on some grounds, since it has a lower DIC, PsML and predictive loss PL than model B.

Therefore model F is a univariate version of model C. It involves separate trivariate normal densities for v_{i1} and v_{i2} with prior means $m_{vm}= \{\kappa_{1m}, \kappa_{2m}, \kappa_{3m}\}$ and covariance matrices Φ_{mv} of order P. Wishart priors for Φ_{mv}^{-1} with identity scale matrix and 3 degrees of freedom are assumed. It is apparent from Table 2 that fit deteriorates under this approach as compared to model C, especially in terms of the DIC; this demonstrates the benefits of a multivariate approach to the case study data.

10 Discussion

This paper has considered spatial variation in the impacts of predictors on a bivariate mortality outcome, and discussed alternative strategies for pooling strength in modelling such variation over several response variables. In particular it has analysed how far male and female suicide variations in England can be explained by spatially varying impacts in social structural factors.

It has been argued and demonstrated that multivariate approaches may be preferable for health outcomes where geographic context suggests some commonality in predictor impacts. Moreover the analysis shows the MMP approach to be competitive with the MCAR prior, while being potentially more versatile in the face of correlated effects (leading to common factor models) or unusual mortality patterns, suggesting departures from normality. The MCAR prior applies most easily to symmetric unimodal normal or student effects; more robust options such as double exponential (Laplace), sometimes applied under a univariate CAR, do not have simple multivariate extensions. By contrast, there is considerable scope for developing non-parametric approaches under the MMP method; for example, Polya tree priors are an alternative option for nonparametric modelling of random effects, and can be set up to put all their probability mass on a continuous distribution (Walker et al, 1999). Another option is the explicit allocation prior of Green & Richardson (2001), where $D_i|K \sim$ Categorical $(p_1,..p_K)$, and $(p_1,..p_K) \sim Dir(\delta)$ with K itself unknown. Alternative baseline densities (e.g. multivariate Student t might be applied), and the scale version of the multivariate t provides area specific weights (Ibrahim & Kleinman, 1999, p 95) that in applications like those here indicate the appropriateness of the model and act to accommodate unusual areas. This avoids the problematic designation of certain areas as 'outliers'.

Whereas disease mapping methods often focus on smoothing to clarify variation in relative risk, the goal of spatially varying regression models is different, namely to identify both a central predictor effect and how that effect is lower or higher in subareas of the study region. Therefore investigation of alternative spatially varying regression effect specifications is not simply a matter of assessing best fit or choosing one model over others. Trying different specifications also amounts to a sensitivity analysis regarding the central impact of the predictors and spatial variation in such impacts. Thus one may ask whether most models identify that a particular risk factor X_k has a 'significant' effect with 95% credible intervals for the average values $\bar{\beta}_{pm}$ entirely positive. The sensitivity analysis also relates to the spatial patterns in impacts: do most models identify the same subareas with distinctly above average or below average effects. Table 1 shows that despite their various discrepant characteristics, all the models report a very similar set of average predictor coefficients. For all models, the effects on male suicide of fragmentation, deprivation and rurality are positive. For females only fragmentation has a clear effect, and outweighs the male coefficient for this area characteristic.

To assess sensitivity with regard to spatial patterning of effects we consider the extremes of the predictor coefficients over the 354 areas for models E and C; model E is the best fitting but with some limitation apparent in its removal of spatial correlation in residuals. There is a clear overlap in these two models' results despite their different structure. Consider the 89 areas in the top quartile on the four sets of coefficients with overall average 'significant' effect, namely male fragmentation, deprivation and rurality coefficients, and female fragmentation coefficients.

The 89 coefficients show an overlap in 59 areas for male fragmentation (i.e. of the 89 areas in the top quartile for β_{i11} coefficients under model C, 59 of these areas appear in the top quartile of β_{i11} coefficients under model E). Figures 2 and 3 confirm this overlap and the correlation is 0.76. For male deprivation and rurality coefficients the overlap in the top quartile is 59/89 and 63/89 while for female fragmentation it is 50/89.

The pattern in Figures 2 and 3 reflects heterogeneity in fragmentation scores between inner and outer city areas (e.g. in London) and between coastal resorts (e.g. on the southern English coast and in SW England) and adjacent non-resort areas. Local differences in fragmentation scores in these parts of the map are linked to suicide risk variations and hence the fragmentation coefficients are high. In other subregions variations between adjacent areas in fragmentation are less and their link to suicide risk is less clear (e.g. in East Anglia).

One possible reason for differences in the density of the coefficients between models E and C is the possibility of extreme effects and non-normality; for example, Figure 4 shows the histogram of male deprivation coefficients under model E, with some evidence of multiple modes.

The analysis here has demonstrated (e.g. via the consistency in findings just mentioned and the consistency in average effects in Table 1) that a varying coefficient analysis in a multivariate mortality model produces sensible results. The alternative and more common methodology of assuming spatially fixed regression effects in spatial epidemiology and "mopping up" residual correlation through a spatially varying intercept may neglect a substantively important feature in the observations.

Of methodological consequence, the present analysis has demonstrated the benefit of allowing for non-normal nonparametric density in the regression effects; model E was closely competitive to model C among the models applied to the suicide data in the paper's case study. So applying the MCAR approach without some critical evaluation may also be missing some important features in the data.

11 References

Anselin, L (2001) Spatial Econometrics. In: A Companion to Theoretical Econometrics (Baltagi, B, ed.),. Oxford: Basil Blackwell, 310-330.

Barnard, J, McCulloch, R, Meng, X. (2000) Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. Statistica Sinica, 10, 1281–1312.

Besag, J, York. J, Mollié, A (1991) Bayesian image restoration with two applications in spatial statistics. Annals of the Institute of Statistics and Mathematics. 43, 1-59.

Best, N, Ickstadt, K, Wolpert, R, Cockings, S, Elliott, P, Bennett, J, Bottle, A, Reed, S (2002) Modeling the impact of traffic-related air pollution on childhood respiratory illness. In: Case Studies in Bayesian Statistics, Volume V (Gatsonis C, Kass R, Carlin B, Carriquiry A, Gelman A, Verdinelli I, West M, eds.), 183-259.

Browne, W, Goldstein, H, Rasbash, J (2001) Multiple membership multiple classification (MMMC) models. Statistical Modelling 1, 103-124.

Chib, S, Winkelmann, R (2001) Markov chain Monte Carlo analysis of correlated count data. Journal of Business & Economic Statistics, 19, 428-435

Congdon, P (1996) Suicide and parasuicide in London: a small area study, Urban Studies, 33, 137-158

Congdon, P (2001) Bayesian Statistical Modelling, John Wiley

Congdon, P (2002) A life table approach to small area health need profiling, Statistical Modelling, 2,1-26

Dorling, D, Gunnell, D (2003) Suicide: the spatial and social components of despair in Britain 1980-2000. Transactions of the Institute of British Geographers, 28, 442-460

Escobar, M, West, M (1995) Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association, 90, 577–588

Ferguson, T (1973) A Bayesian analysis of some nonparametric problems. The Annals of Statistics, 1, 209–230.

Ferguson, T (1974) Prior distributions on spaces of probability measures. The Annals of Statistics, 2, 615–629.

Fotheringham, A, Charlton, M, Brunsdon, C (2003) Geographically Weighted Regression. John Wiley: Chichester

Gamerman, D, Moreira, A, Rue, H (2003) Space-varying regression models: specifications and simulation, Comp. Statist. & Data Analysis, 42, 513-533

Gelfand, A (1996) Model determination using sampling-based methods. In: Markov Chain Monte Carlo in Practice (Gilks, W, Richardson, S, Spiegelhalter, D, eds.) Chapman Hall, London, 145-161.

Gelfand, A, Ghosh, S (1998) Model choice: a minimum posterior predictive loss approach, Biometrika, 85, 1-11

Gelfand, A, Kottas, A, MacEachern, S (2004) Bayesian nonparametric spatial modeling with Dirichlet processes mixing, Manuscript, ISDS, Duke University, Durham, NC

Gelman, A, Carlin, J, Stern, H, Rubin, D (2003) Bayesian Data Analysis, 2nd Edition, CRC Press

Green, P, Richardson, S (2001) Modelling heterogeneity with and without the Dirichlet process. Scandinavian Journal of Statistics, 28, 355-377

Gunnell, D, Peters, T, Kammerling, R, Brooks, J (1995) Relation between parasuicide, suicide, psychiatric admissions and socio-economic deprivation. British Medical Journal, 311, 226-230.

Hawton, K, Fagg, J, Simkin, S, Harriss, L, Malmberg, A, Smith, D (1999) The geographical distribution of suicides by farmers in England and Wales. Social Psychiatry and Psychiatric Epidemiology, 34, 122-127

Hogan, J, Tchernis, R (2004) Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. J Amer Stat Assn, 99, 314-324

Ibrahim, J, Kleinman, K (1999) Semiparametric Bayesian methods for random effects models. In Practical Nonparametric and Semiparametric Bayesian Statistics (D. Dey, P. Mueller, D. Sinha, eds.). Springer Lecture Notes.

Ishwaran, H, Zarepour, M (2002a) Exact and approximate sum-representations for the Dirichlet process. Can. J. Statist. 30, 269-283.

Ishwaran, H, Zarepour, M (2002b) Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. J. Comp. Graph. Statist. 11, 508-532

Johnson, V, Albert, J (1999) Ordinal Data Modeling. Springer: New York.

Langford, I, Leyland, A, Rasbash, J, Goldstein, H (1999) Multilevel modelling of the geographical distributions of diseases. Appl. Statist., 48, 253-268.

Lawson, A, Clark, A. (2002) Spatial mixture relative risk models applied to disease mapping. Statistics in Medicine, 21, 359-370

Leyland, A, Langford, I, Rasbash, J, Goldstein, H (2000) Multivariate spatial models for event data. Statistics in Medicine, 19, 2469-2478

Marshall, E, Spiegelhalter, D (2003) Approximate cross-validatory predictive checks in disease mapping models, Statistics In Medicine, 22, 1649–1660

Middleton, N, Gunnell, D, Frankel, S, Whitley, E, Dorling, D (2003) Urban–rural differences in suicide trends in young adults: England and Wales, 1981–1998. Social Science & Medicine, 57, 1183–1194

Mukhopadhyay, S, Gelfand, A (1997) Dirichlet process mixed generalized linear models. J. Amer. Statist. Assoc., 92, 633-639

Mueller, P, Quintana, F (2004) Nonparametric Bayesian data analysis. Statistical Science, 19, 95-110

Pretorius, A, Van Der Merwe, A (2002) Non-parametric Bayesian Estimation for the Mixed Linear Model. In Bayesian Methods with Applications to Science, Policy and Official Statistics, Selected Papers from ISBA, 2001 Monographs of Official Statistics, George, I (ed). Eurostat, 421–427

Sethuraman, J (1994) A constructive definition of Dirichlet priors. Statistica Sinica, 4, 639-650.

Spiegelhalter, D, Best, N, Carlin, B, van der Linde, A (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, 64B, 583-639.

Spiegelhalter, D, Thomas, A, Best, N, Lunn, D (2003) WinBUGS User Manual. Version 1.4 (http://www.mrc-bsu.cam.ac.uk/bugs)

Walker, S, Damien, P, Laud, P, Smith, A (1999) Bayesian nonparametric inference for distributions and related functions. J. Royal Statist. Soc. B, 61, 485-527.

Wang, F, Wall, M (2003) Generalized common spatial factor model. Biostatistics, 569-582

West, M, Turner, D (1994) Deconvolution of mixtures in analysis of neural synaptic transmission. The Statistician, 43, 31-43

West, M, Mueller, P, Escobar, D (1994) Hierarchical priors and mixture models, with application in regression and density estimation. In Freeman, P, Smith, A (eds), Aspects of Uncertainty: A Tribute to D.V. Lindley, 363-386.

Whitley E, Gunnell D, Dorling D, Smith G (1999) Ecological study of social fragmentation, poverty, and suicide. British Medical Journal, 319, 1034-7.

Table 1 Mean Regression Effects (Spatial Regression Heterogeneity)

	Regression Coefficient (and Moran Index)	Mean	2.5%	97.5%
Model A	Male, Fragmentation	0.096	0.078	0.113
	Male, Deprivation	0.050	0.033	0.066
	Male, Rurality	0.077	0.057	0.098
	Female, Fragmentation	0.151	0.123	0.175
	Female, Deprivation	0.008	-0.017	0.035
	Female, Rurality	0.019	-0.016	0.053
	Moran (Males)	0.098	0.088	0.114
	Moran (Females)	0.134	0.128	0.146
Model B	Male, Fragmentation	0.085	0.047	0.123
	Male, Deprivation	0.041	0.004	0.079
	Male, Rurality	0.061	0.023	0.100
	Female, Fragmentation	0.136	0.084	0.188
	Female, Deprivation	0.005	-0.042	0.053
	Female, Rurality	-0.023	-0.070	0.035
	Moran (Males)	0.025	-0.038	0.090
	Moran (Females)	0.033	-0.023	0.091
Model C	Male, Fragmentation	0.085	0.039	0.131
	Male, Deprivation	0.059	0.015	0.101
	Male, Rurality	0.067	0.023	0.111
	Female, Fragmentation	0.148	0.083	0.211
	Female, Deprivation	0.014	-0.038	0.071
	Female, Rurality	-0.007	-0.070	0.059
	Moran (Males)	0.019	-0.041	0.080
	Moran (Females)	0.045	-0.014	0.103
Model D1	Male, Fragmentation	0.085	0.039	0.134
	Male, Deprivation	0.061	0.019	0.106
	Male, Rurality	0.070	0.028	0.113
	Female, Fragmentation	0.151	0.086	0.211
	Female, Deprivation	0.012	-0.042	0.072
	Female, Rurality	-0.012	-0.073	0.049
	Moran (Males)	0.021	-0.038	0.082
	Moran (Females)	0.044	-0.009	0.101
Model E	Male, Fragmentation	0.096	0.067	0.122
	Male, Deprivation	0.058	0.033	0.081
	Male, Rurality	0.072	0.048	0.097
	Female, Fragmentation	0.157	0.116	0.194
	Female, Deprivation	0.017	-0.017	0.055
	Female, Rurality	0.008	-0.043	0.053
	Moran (Males)	0.031	-0.015	0.073
	Moran (Females)	0.050	0.005	0.100

Table 2 Model Fit and Checking Criteria

Model		$Dev(\bar{\theta})$	DIC	d_e	PsML	Predictive Loss (c=1)	Outliers under Mixed Predictive sampling
A	Fixed Coefficients	4668	4684	8	-2345	48570	80
B	MCAR	3983	4573	295	-2314	44295	18
C	Multiple membership Fully Parametric Multivariate Prior	4043	4556	257	-2295	44084	26
D1	MMP Factor Model (Predictor Specific Factors)	4198	4548	175	-2303	43725	35
D2	MMP Factor Model (Outcome Specific Factors)	4288	4602	157	-2336	45619	48
E	Multiple membership NonParametric Multivariate Prior (κ=0.5)	4276	4546	135	-2298	43785	4
F	Multiple membership Fully Parametric Univariate Priors	4298	4854	278	-2313	44434	24

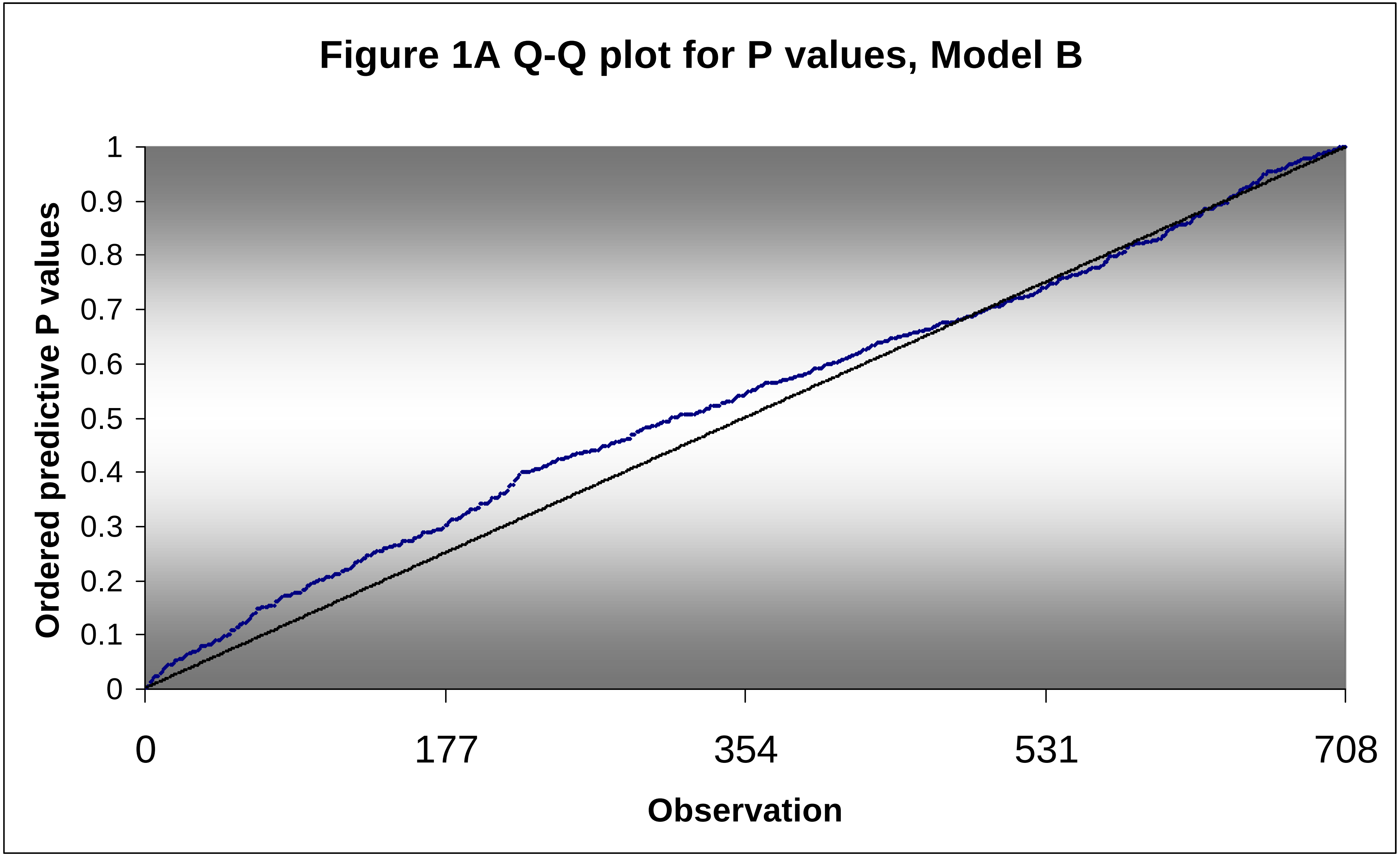


Figure 1A Q-Q plot for P values, Model B
Ordered predictive P values
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
177
354
531
708
Observation

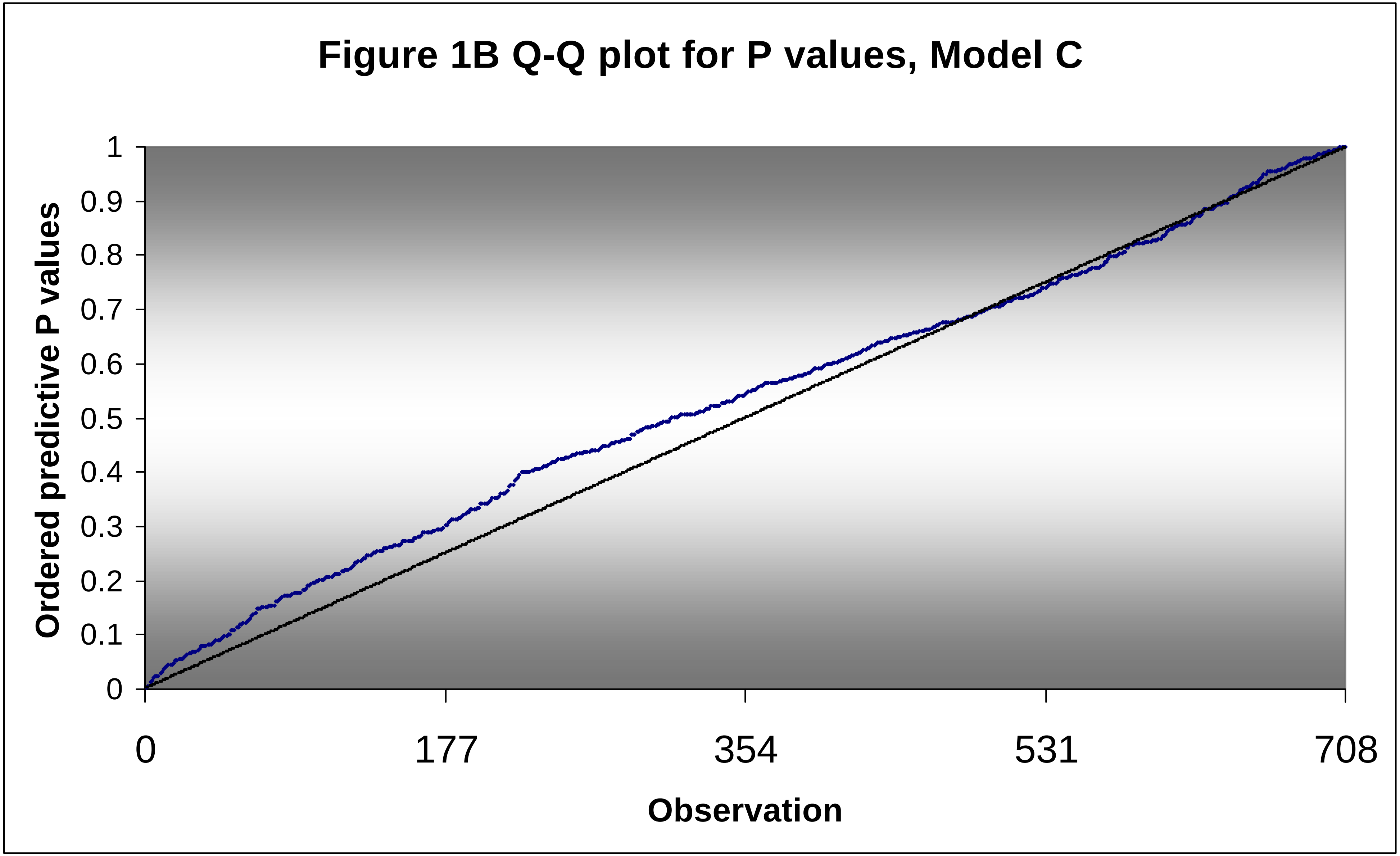


Figure 1B Q-Q plot for P values, Model C
Ordered predictive P values
Observation
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
177
354
531
708

Figure 2 Fragmentation Coefficients (M), Model E

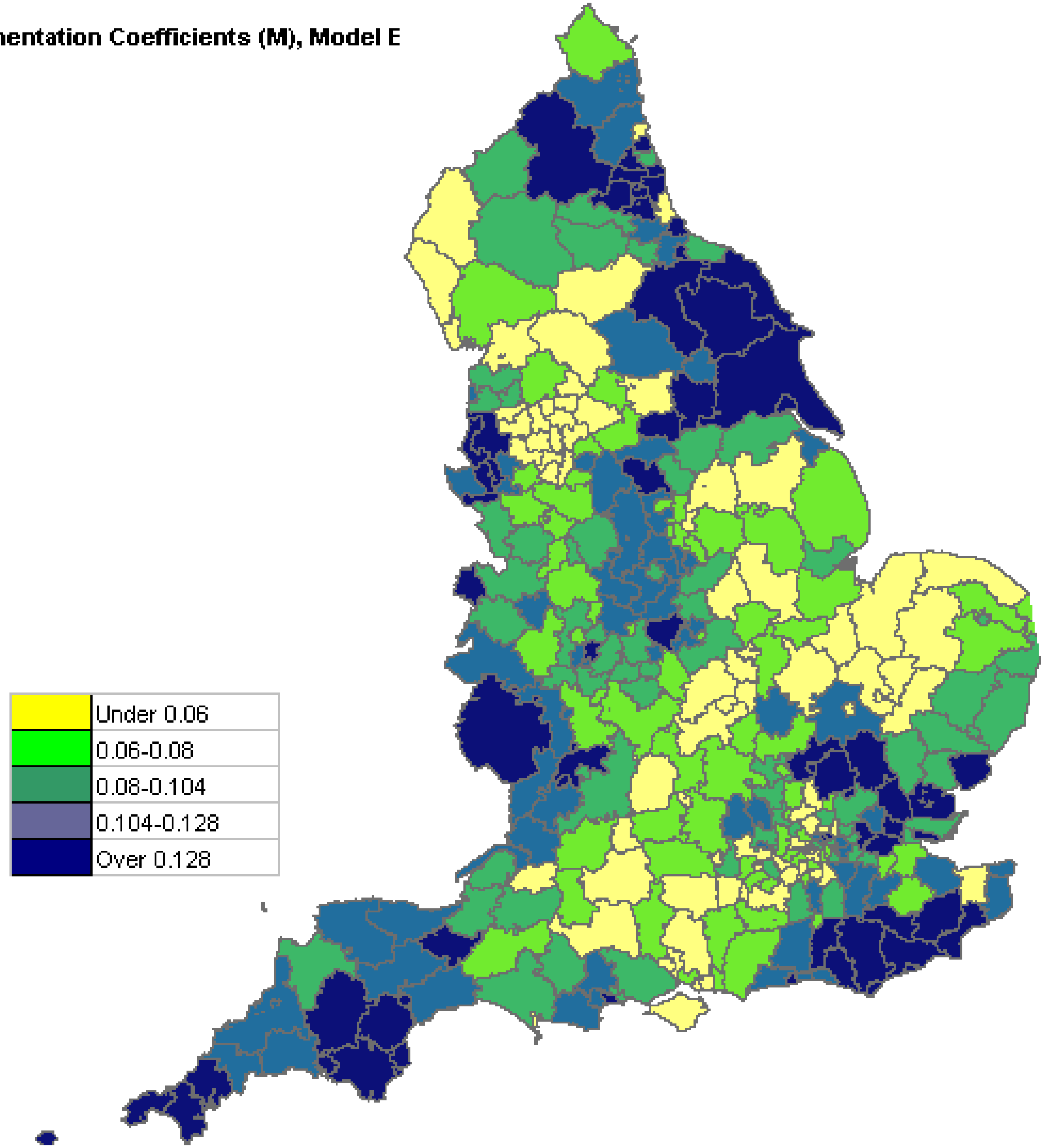
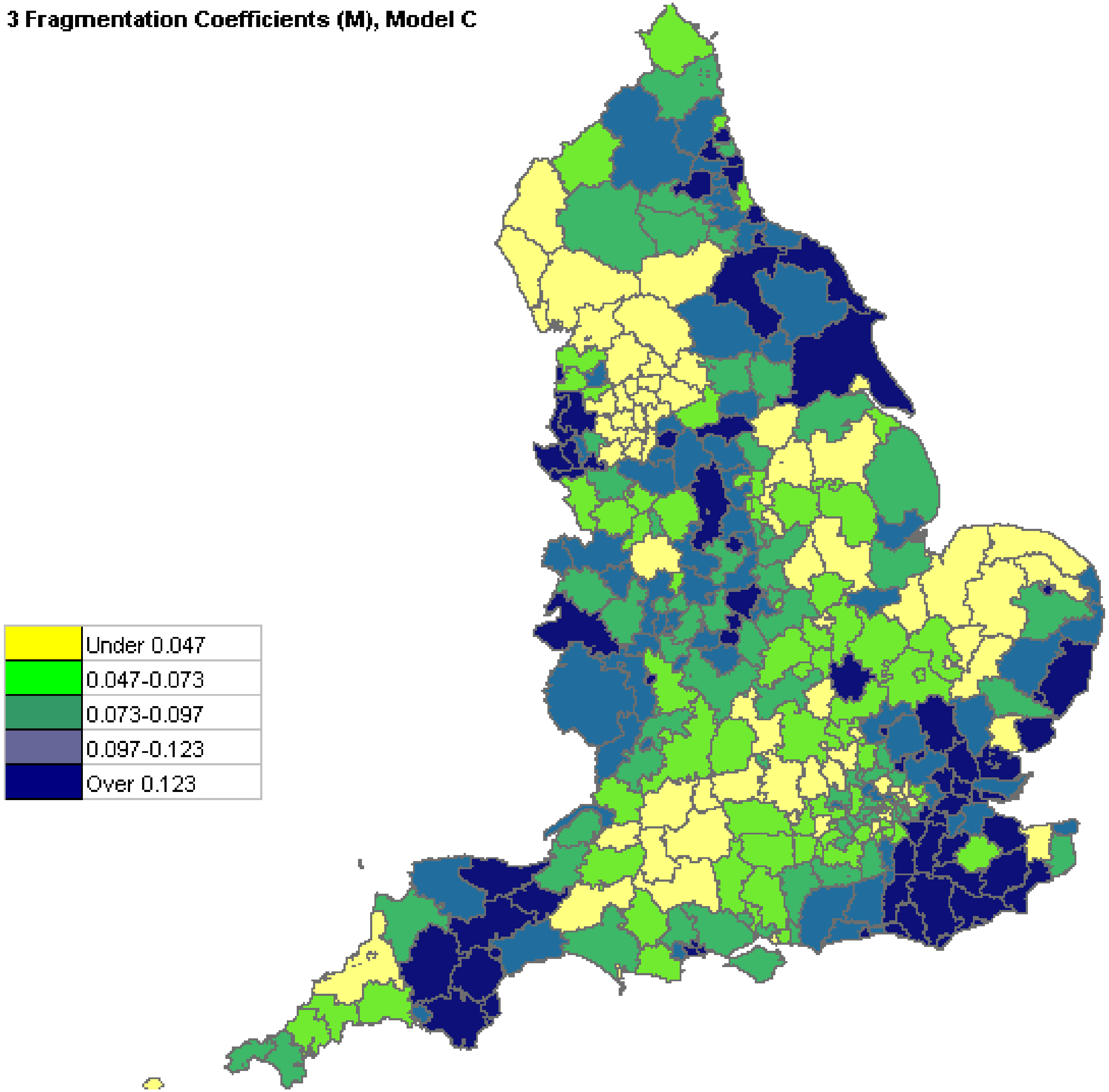


Figure 3 Fragmentation Coefficients (M), Model C



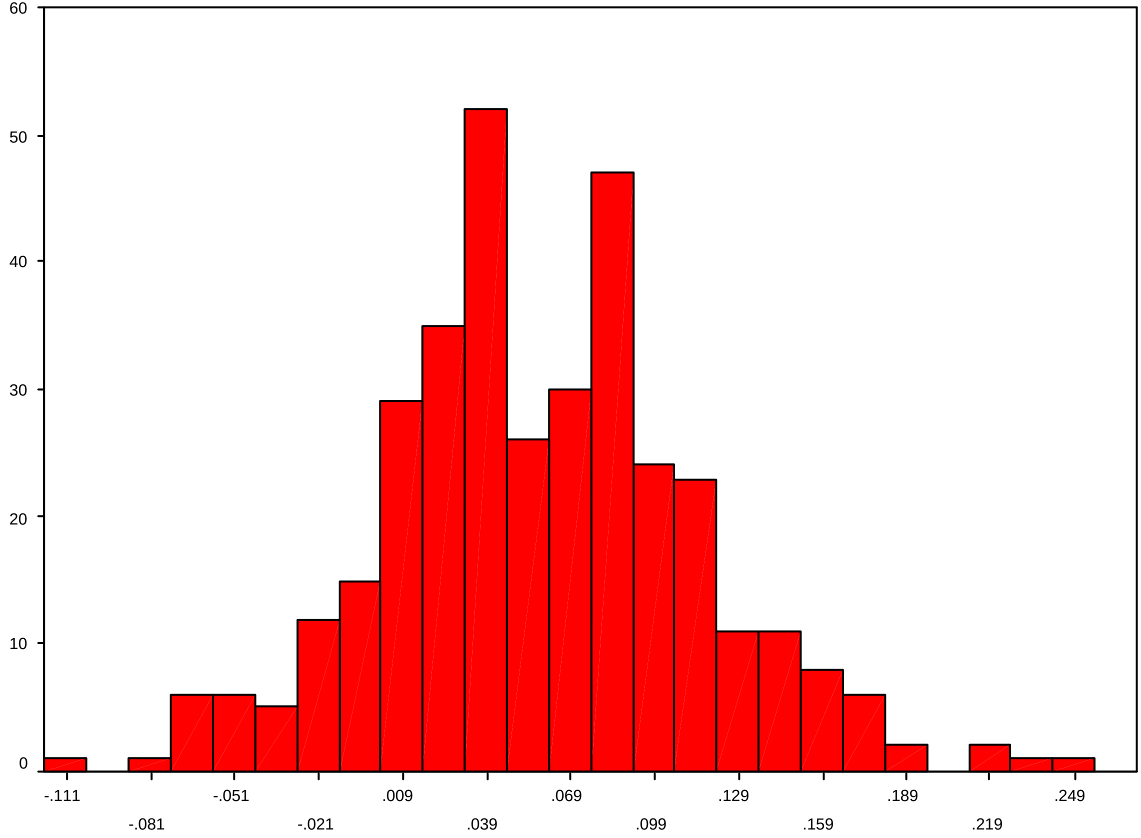


Figure 4 Male Deprivation Coefficients, Model E