Dipartimento di Statistica
"Giuseppe Parenti"

W O R K I N G P A P E R 2012/06

Disease Mapping via Negative Binomial M-quantile regression

Ray Chambers,
Emanuela Dreassi, Nicola Salvati

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Disease Mapping via Negative Binomial M-quantile regression

Ray Chambers†, Emanuela Dreassi‡, Nicola Salvati§

† Centre for Statistical and Survey Methodology, University of Wollongong, New South Wales 2522, Australia
ray@uow.edu.au

‡Dipartimento di Statistica "G. Parenti", Università degli Studi di Firenze, Viale Morgagni, 59 - I 50134 Florence, Italy
deassi@ds.unifi.it

§Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa, Via Ridolfi, 10 - I 56124 Pisa, Italy
salvati@ec.unipi.it

Abstract

A new approach to ecological regression on disease mapping is introduced: a semi-parametric method based on M-quantile models. We define a Negative Binomial M-quantile model as an alternative to Empirical Bayes or fully Bayesian approaches to disease mapping. The proposed method is easily made robust against outlying data values for covariates. Robust ecological regression on disease mapping is desirable since covariates at area level usually present measure-type error. Differences between M-quantile and usual random effects models are discussed and the alternative approaches are compared using the Scottish Lip cancer example and a simulation experiment. The example shows that the Negative Binomial M-quantile model confirms results obtained by other methods, but it seems to have less shrinkage effect than the Empirical Bayes method, so reducing the problem of oversmoothing. The simulation experiment suggests that the new model presents smaller root mean square error. The Negative Binomial M-quantile is also extended to accounting for spatial structure between areas following a Geographically Weighted Regression strategy.

Keywords: Ecological regression; Overdispersed count data; Robust models; Spatial correlation.

1 Introduction

Disease mapping involves the analysis of disease incidence or mortality data often available as aggregate counts over a geographical region subdivided for administrative purposes. Such data are often relatively easy to be obtained from government sources. More difficult is to obtain data, at aggregated level,

on explanatory covariates that could be considered as known or putative risk factors.

Ecological regression on disease mapping mainly focuses on the estimation of risk in administrative regions and on the analysis of the association between risk factors and disease. In ecological analysis related to disease mapping, data usually exhibit overdispersion. Hence, Clayton and Kaldor (1987) proposed the use of a Poisson-gamma model for relative risks using an Empirical Bayes approach (referred to as EB below). This model was generalized by Besag *and others* (1991) into a fully Bayesian setting using a Hierarchical Bayesian model with or without a spatial structure (hereafter BYM). Ecological disease mapping typically relies on regression models that use both covariates and random effects to explain variation between areas and to take the overdispersion into account. These models depend on strong distributional assumptions and require a formal specification of the random part of the model. On several real examples, the use of spatial area data requires more flexible forms than the usual linear predictor for modelling the dependence of responses on covariates (see, for example, space varying coefficients models: Assuncao (2003)). Moreover, the standard models do not easily allow for outlier-robust inference because of covariates at area level that could be measure-type error prone (i.e. MacNab (2009); MacNab (2010); Wakefield (2007)).

Ecological regression on disease mapping can be regarded as a special case of application of small area methodology Rao (2003) (Chapter 9). The EB method provides reliable estimators of risk by borrowing strength across areas. It belongs to the family of predictors obtained by fitting generalized linear mixed models. EB is applicable to different models, ranging from models for binary or count data to normal linear mixed models. In the latter case, EB and Empirical Best Linear Unbiased Predictor estimators coincide Rao (2003) (Chapter 9). In the case of a continuous response variable, Chambers and Tzavidis (2006) proposed an approach based on M-quantile regression to small area estimation that controls for the effect of outliers and relaxes some of the conventional assumptions on the model. This approach requests weaker parametric assumptions while the use of M-estimation guarantees outlier robust estimation. For these reasons, Chambers *and others* (2012) proposed a new approach to small area estimation for discrete data based on a M-quantile model extending the robust version of the estimating equations for generalized linear models by Cantoni and Ronchetti (2001) to the M-quantile case.

In this paper, we extend the method by Chambers *and others* (2012) to the case of Negative Binomial M-quantile regression (referred to as NBMQ below) for the ecological disease mapping. Roughly speaking, the underlying idea is to model quantiles like parameters of the conditional distribution of the target variable given the covariates. Unlike usual random effects models, NBMQ models do not depend on strong distributional assumptions and are robust to the presence of outliers due to measure-type error on covariates.

In disease mapping, data are usually spatially structured and the model should include a suitable spatial component to take this fact into account. In the NBMQ models introduced in this paper, the spatial structure is captured by

appropriate weights at the estimation step (see Salvati *and others* (2012)) using a Geographically Weighting Regression philosophy (referred to as NBMQGWR below).

Negative Binomial M-quantile and usual random effects models are compared using the Scottish Lip cancer example and a simulation experiment. The example shows that the Negative Binomial M-quantile model confirms results obtained by other methods, but it seems to have less shrinkage effect than the Empirical Bayes method, so reducing the problem of oversmoothing. The inclusion of the spatial structure in the model gives results very similar to Besag *and others* (1991) spatial model. The simulation experiment suggests that the new model presents smaller root mean square error.

This paper is organized as follows. In Section 2, the Negative Binomial model to describe overdispersed count data and disease mapping is reviewed. In Section 3, the Negative Binomial robust model, extending the class of models introduced by Cantoni and Ronchetti (2001), is introduced. In Section 4, the Negative Binomial M-quantile model for overdispersed count data is proposed and applied in disease mapping. Moreover, in the same Section, we propose a nonparametric bootstrap method for estimating the MSE, that is easy to implement by extending existing approach by Chambers *and others* (2012). In Section 5, differences between NBMQ and random effects models such as EB and BYM are discussed and compared using the Scottish Lip cancer example. In Section 6, for comparing bias and root mean squared error of the considered models, a simulation study is conducted. In Section 7, the NBMQGWR is introduced and it is compared with the Besag *and others* (1991) when a set of spatially structured random terms are considered. Conclusions are reported in Section 8.

2 Overdispersed count data

Usually, the Poisson model is useful for describing the mean but underestimates the variance of the data. The are essentially three ways for dealing with this fact. One is to use the same estimating function for the mean, but to base inference on the more robust sandwich covariance matrix estimator. The second is to use a Quasi-Poisson model. The third is modeling overdispersed count data by a Negative Binomial distribution which can arise as a Gamma mixture of Poisson distributions. This paper focus on the latter way.

Let $Y \sim \text{Poisson}(\lambda)$ and $\lambda \sim \text{Gamma}(\theta, \alpha)$. The compound model is a Negative Binomial distribution

$$p(y;\theta,\alpha) = \left(\begin{array}{c} y+\theta-1 \\ \theta-1 \end{array} \right) \left(\frac{\alpha}{1+\alpha} \right)^{\theta} \left(\frac{1}{1+\alpha} \right)^{y}$$

for $y = 0, 1, 2, \ldots$ (number of failure to obtain θ success), with $p = \alpha/(1+\alpha)$ the success probability. We obtain $\text{E}[Y] = \theta/\alpha$ and $\text{Var}[Y] = \theta/\alpha + \theta/\alpha^2$, so that θ/α^2 represent the Poisson overdispersion. Parameterizing according the

mean value $\mu = \theta/\alpha$, one obtains $\alpha = \theta/\mu$ and

$$p(y;\mu,\theta) = \frac{\Gamma(y+\theta)}{\Gamma(\theta)y!}\left(\frac{\theta}{\mu+\theta}\right)^{\theta}\left(\frac{\mu}{\mu+\theta}\right)^{y}$$

where now $E[Y] = \mu$, $\text{Var}[Y] = V[\mu] = \mu + \frac{\mu^2}{\theta}$. It must be noted that the variance is now equal to the Poisson variance μ plus the extra-variance component. Because the variance is a quadratic function of the mean, this model is referred to as the NEGBIN2 or NB2 model in Cameron and Trivedi (1998). The value $1/\theta$ is directly related to the amount of overdispersion in the data: increasing values of $1/\theta$ suggest increasing amounts of overdispersion. For every fixed θ, Negative Binomial distribution is a member of the exponential family.

Since our interest is in ecological regression, when a log-linear specification is used, Y represents the response variable and $\mathbf{x}$ a $p \times 1$ vector of explanatory variables (including the constant). A Poisson model would stipulate that the distribution of Y_i given $\mathbf{x}_i$ is Poisson with mean equal to $\mu(\mathbf{x}_i^T) = \exp\eta_i = \exp(\mathbf{x}_i^T\boldsymbol{\beta})$, with $\boldsymbol{\beta}$ is a vector of p regression parameters. Similarly, Negative Binomial regression model, where the link function is $\log(\cdot)$ to easily compare with Poisson, considers as mean parameter $\mu(\mathbf{x}_i) = \exp\eta_i = \exp(\mathbf{x}_i^T\boldsymbol{\beta})$. Considering n observations we have $\log(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$. This is a special case of the Generalized Linear Model (GLM), while the Negative Binomial model is an exponential family for θ fixed but not in general. However, in line with a standard practice (McCullagh and Nelder (1989); Breslow (1984); Lawless (1987)), a GLM methodology can be used as well, after replacing θ with a suitable estimate $\hat{\theta}$ (obtained, e.g., using the method of moments) and by iterating estimation of $\boldsymbol{\beta}$ given $\hat{\theta}$.

Log-linear models for count data represent the basic models to estimate relative risks of mortality when a set of deaths counts are available at aggregate level on a map. In the next section, the 'standard' methods used for disease mapping are reviewed. These methods will be the cornerstone to evaluate the performances of our approach.

2.1 Models for disease mapping

Consider a region partitioned into n areas. Let y_i denote the observed number of deaths for area $i = 1,\ldots,n$. Each y_i is assumed to be a realization of a random variable Y_i, where $Y_1,\ldots,Y_n$ are independent with $Y_i \sim \text{Poisson}(\mu_i)$. Here, $\mu_i = E_i\lambda_i$ where E_i represents the expected cases in area i-th and λ_i the relative risk. This is the basic model, when no covariates are considered. The likelihood function for the entire data is the corresponding product of Poisson terms. The MLE estimates for λ_i is $\text{SMR}_i = y_i/E_i$, the so called *Standardized Mortality Rate*. Since this type of data typically exhibits substantial overdispersion, James-Stein type estimators are preferred (see Efron and Morris (1973)). Following Clayton and Kaldor (1987) the λ_i are assumed independently and identically distributed as a Gamma(θ,α). The compound model is a Negative Binomial model with mean $\theta(E_i/\alpha)$ and variance $\theta(E_i/\alpha) + \theta(E_i/\alpha)^2$. Each

λ_i, conditionally to the others parameters and data, has a posteriori Gamma distribution with mean $\mathrm{E}[\lambda_i \mid y_i, \theta, \alpha] = (y_i + \theta)/(E_i + \alpha)$. This is the empirical Bayes estimate once the parameters α and θ are replaced by their estimates (using the method of moments or maximum likelihood estimation). These values could be considered as a weighed mean between SMR and the prior mean for λ_i, with the weights depend on E_i. We can easily include into the model a set of covariates to perform an ecological regression.

The Empirical Bayes method has been extended to a fully Bayesian one by Besag *and others* (1991). Following their standard model

$$\log(\lambda_i) = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij} + u_i + v_i$$

where β_0 represents an intercept, such as an overall risk level; $\beta_1, \ldots, \beta_{p-1}$ is a set of coefficients; u_i is a spatially structured random effect (called *clustering*) and v_i a spatially unstructured (called *heterogeneity*) random effect. The prior distributions for the model parameters are as follows. The intercept β_0 is given a flat non-informative distribution. The coefficients β_j are given an uninformative normal distribution with mean zero. The heterogeneity terms v_i are independent, each v_i being Normal with mean 0 and variance ψ_v^{-1}, where ψ_v represents the precision parameter. The clustering terms u_i, using Gaussian Markov random fields (GMRFs) models in order to cope the spatial structure, are modeled conditionally on $u_{l\sim i}$ terms, as Normal $(\bar{u_i}, (\lambda_u n_i)^{-1})$ where $\bar{u_i} = \sum_{l\sim i} \frac{u_l}{n_i}$. Here $l \sim i$ $(l = 1, \ldots, n)$ indicates adjacent areas to i-th ones (adjacent means that two areas share an edge) and n_i represents their number. The hyperprior distributions of the precision parameters ψ_v and ψ_u are assumed to be Gamma $(0.5, 0.0005)$ as suggested by Kensall and Wakefield (1999). The marginal posterior distributions of the parameters of interest are approximated by Monte Carlo Markov Chain methods. The model could be considered without any spatially structured random effects u_i (named BYM) or considering these (hereafter BYMspatial) to take into account for the spatial characteristic of the data.

3 Robust estimation for Negative Binomial model

Cantoni and Ronchetti (2001) propose a robust inference for generalized linear models based on quasi-likelihood. They consider a general class of M-estimators of Mallows's type, where the influence of deviations on y and on $\mathbf{X}$ are bounded separately. The robust version of the generalized linear model estimating equations is

$$n^{-1} \sum_{i=1}^{n} \boldsymbol{\psi}(y_i, \mu_i) = \mathbf{0} \tag{1}$$

where $\boldsymbol{\psi}(y_i, \mu_i) = v(y_i, \mu_i) w(\mathbf{x}_i) \mu_i' - a(\boldsymbol{\beta})$, $E[Y_i] = \mu_i$, $V[Y_i] = V(\mu_i)$, $\mu_i = \mu_i(\boldsymbol{\beta}) = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$, μ_i' is its derivative and $a(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} E[v(y_i, \mu_i)] w(\mathbf{x}_i) \mu_i'$

ensures the Fisher consistency of the estimator. The bounded $v(y, \mu)$ function is introduced to control deviation in y-space, whereas weights $w(\mathbf{X})$ are used to down-weight the leverage points. When $w(\mathbf{x}_i) = 1 \; \forall \; i$ Cantoni and Ronchetti (2001) call the estimator the Huber quasi-likelihood estimator. The authors present the robust estimation for Binomial and Poisson models by using the Pearson residuals and the ψ_c Huber function. The solution of the estimating equations (1) can be obtained numerically by a Fisher scoring procedure.

In this Section we extend the robust estimation to the Negative Binomial model. This model, when parametrized by the mean, with the parameter θ fixed, is an exponential family (see Cameron and Trivedi (1998)). Under Negative Binomial model we use the estimating equations

$$\Psi(\boldsymbol{\beta}) := n^{-1} \sum_{i=1}^{n} \boldsymbol{\psi}_c(y_i, \mu_i) = \mathbf{0} \tag{2}$$

where $\boldsymbol{\psi}_c(y_i, \mu_i) = \left\{ \psi_c(r_i) w(\mathbf{x}_i) \frac{1}{V^{1/2}(\mu_i)} \mu_i' - a(\boldsymbol{\beta}) \right\}$, $r_i = \frac{y_i - \mu_i}{V^{1/2}(\mu_i)}$ are the Pearson residuals, ψ_c is the Huber Proposal 2 influence function, $\psi_c(r) = c\, I(-c < r < c) + c\, \text{sgn}(r)\, I(|r| \geqslant c)$, $\mu_i = t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, t_i is the offset term, $\mu_i' = \mu_i \mathbf{x}_i^T$, $V(\mu_i) = \mu_i + \frac{\mu_i^2}{\theta}$ and $\theta > 0$ is a shape parameter. The correction term $a(\boldsymbol{\beta}) = 1/n \sum_{i=1}^{n} E[\psi_c(r_i)] V^{-1/2}(\mu_i) \mu_i'$ can be computed explicitly for the Negative Binomial model, as shown in Appendix. The parameter θ has to be estimated by using a robust method to maintain the robustness properties gained in the estimation of $\boldsymbol{\beta}$. We propose the robust scale Huber's Proposal 2 estimator Huber (1981) defined by

$$n^{-1} \sum_{i=1}^{n} \left\{ \psi_c^2(r_i) - E\left[\psi_c^2 \left(\frac{Y_i - \mu_i}{V^{1/2}(\mu_i)} \right) \right] \right\} = \mathbf{0}, \tag{3}$$

where $E\left[\psi_c^2 \left(\frac{Y_i - \mu_i}{V^{1/2}(\mu_i)} \right) \right]$ is a constant that ensures Fisher consistency for the estimation of θ (see the Appendix for its computation) and ψ_c can be chosen as in (2). The equations (2) and (3) have to be solved simultaneously, but in practice a two-step procedure is often used: (i) starting from a first guess for θ, an estimate of $\boldsymbol{\beta}$ is obtained, which in turn is used in (3), and so on until convergence; (ii) estimating θ by using residuals of robust Quasi-Poisson model and then, given this estimate, $\boldsymbol{\beta}$ is obtained by solving (2).

Following Cantoni and Ronchetti (2001) for estimating the variance of the estimated regression coefficients $\hat{\boldsymbol{\beta}}$, assuming that $\psi_c(\cdot)$ is a bounded and non-decreasing function, we can write down a sandwich-type estimator as

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{W}^{-1} \mathbf{V} (\mathbf{W}^T)^{-1}. \tag{4}$$

In expression (4) the matrices $\mathbf{W}$ and $\mathbf{V}$ can be computed for the Mallows quasi-likelihood estimator:

$$\mathbf{V} = \frac{1}{n} \mathbf{X}^T \mathbf{D} \mathbf{X} - a(\boldsymbol{\beta}) a(\boldsymbol{\beta})^T, \tag{5}$$

where $\mathbf{D}$ is a diagonal matrix with elements $d_i = E[\psi_c^2(r_i)]w^2(\mathbf{x}_i)\frac{1}{V(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$ with $\eta_i = g(\mathbf{x}_i^T\boldsymbol{\beta}) = \mathbf{x}_i^T\boldsymbol{\beta}$, and

$$\mathbf{W} = \frac{1}{n}\mathbf{X}^T\mathbf{B}\mathbf{X}, \tag{6}$$

where $\mathbf{B}$ is a diagonal matrix with elements $b_i = E[\psi_c(r_i)\left(\frac{\partial}{\partial \mu_i}\right)\log h(y_i|\mathbf{x}_i,\mu_i)]\frac{1}{V^{1/2}(\mu_i)}w(\mathbf{x}_i)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$ with $h(\cdot)$ is the conditional density of $y_i|\mathbf{x}_i$ and $\frac{\partial \log(h(y_i;\theta,\mu_i))}{\partial \mu_i} = \sum_{i=1}^n \frac{y_i-\mu_i}{V(\mu_i)}$ and the elements of $\mathbf{D}$ and $\mathbf{B}$ are computed in Appendix. An estimator of the first order approximation (4) is then

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{W}}^{-1}\hat{\mathbf{V}}(\hat{\mathbf{W}}^T)^{-1}. \tag{7}$$

4 Negative Binomial M-quantile regression

We define an extension of linear M-quantile regression to overdispersed count data. To begin with, the M-quantile regression (Breckling and Chambers (1988)) is a 'quantile-like' generalization of regression based on the influence function (M-regression). The relationship between sample M-quantiles and standard M-estimates of a regression function is the same as that between sample quantiles and sample median. In fact, the M-quantile regression line of order q, $q \in (0,1)$, of a random variable Y with continuous distribution function $F(\cdot)$ is defined as the solution $Q_q(\mathbf{X};\psi) = \mathbf{X}\boldsymbol{\beta}_q$ to

$$E\left[\psi_q\left(\frac{Y - Q_q(\mathbf{X};\psi)}{\sigma_q}\right)\right] = 0, \tag{8}$$

where σ_q is a suitable measure of the scale of the random variable $Y - Q_q(\mathbf{X};\psi)$, $\psi_q(r) = 2\psi_c(r/\sigma_q)\left[q\,I(r>0) + (1-q)I(r\leqslant 0)\right]$ and ψ_c is an appropriately chosen influence function. Here $\boldsymbol{\beta}_q$ is the $p\times 1$ vector of the regression coefficients at quantile qth. The general M-estimator of $\boldsymbol{\beta}_q$ can be obtained by solving the set of estimating equations

$$n^{-1}\sum_{i=1}^n \psi_q(r_{iq})\mathbf{x}_i = \mathbf{0}, \tag{9}$$

with respect to $\boldsymbol{\beta}_q$ with $r_{iq} = y_i - \mathbf{x}_i^T\boldsymbol{\beta}_q$ and σ_q is estimated by s, a robust estimate of scale, e.g. the median absolute deviation estimate $s = \mathrm{median}|r_{iq}|/0.6745$. Being a robust regression model, it can be fitted using an IRLS algorithm that guarantees the convergence to a unique solution Kokic *and others* (1997).

There are no agreed definitions of an M-quantile regression function when Y is overdispersed count data (rates parameterized). The most appealing, of course, is using a log-linear specification under the Negative Binomial model

$$Q_q(\mathbf{X};\psi) = \mathbf{t}\exp(\boldsymbol{\eta}_q), \tag{10}$$

where $\boldsymbol{\eta}_q = \mathbf{X}\boldsymbol{\beta}_q$ is the linear predictor and $\mathbf{t}$ is the vector of offset terms (expected cases of death).

We consider the extensions of (1) to the M-quantile case. Under the M-quantile framework the estimating equations can be written as

$$\Psi(\boldsymbol{\beta}_q) := n^{-1}\sum_{i=1}^{n}\boldsymbol{\psi}_q(y_i, Q_q(\mathbf{x}_i;\psi)) = \mathbf{0}, \tag{11}$$

where $\boldsymbol{\psi}_q(y_i, Q_q(\mathbf{x}_i;\psi)) = \left[\psi_q(r_{iq})w(\mathbf{x}_i)\frac{Q'_q(\mathbf{x}_i;\psi)}{V^{1/2}(Q_q(\mathbf{x}_i;\psi))} - a(\boldsymbol{\beta}_q)\right]$, $r_{iq} = \frac{y_i - Q_q(\mathbf{x}_i;\psi)}{V^{1/2}(Q_q(\mathbf{x}_i;\psi))}$, $V(Q_q(\mathbf{x}_i;\psi)) = Q_q(\mathbf{x}_i;\psi) + \frac{Q_q(\mathbf{x}_i;\psi)^2}{\theta_q}$, $\theta_q > 0$ is a shape parameter, $Q'_q(\mathbf{x}_i;\psi) = Q_q(\mathbf{x}_i;\psi)\mathbf{x}_i$. In addition, by using the results in the Appendix for robust NEGBIN2,

$$\begin{aligned} a(\boldsymbol{\beta}_q) &= 2w_q(r_{iq})\left\{-c\,P\left(Y_i \leqslant j_1\right) + c\,P\left(Y_i \geqslant j_2+1\right) + \right. \\ &+ \left. \frac{Q_q(\mathbf{x}_i;\psi)}{V^{1/2}(Q_q(\mathbf{x}_i;\psi))}P(Y_i = j_1)\left(1+\frac{j_1}{\theta_q}\right) - \frac{Q_q(\mathbf{x}_i;\psi)}{V^{1/2}(Q_q(\mathbf{x}_i;\psi))}P(Y_i = j_2)\left(1+\frac{j_2}{\theta_q}\right)\right\} \end{aligned}$$

where $j_1 = \lfloor Q_q(\mathbf{x}_i;\psi) - cV^{1/2}(Q_q(\mathbf{x}_i;\psi))\rfloor$, $j_2 = \lfloor Q_q(\mathbf{x}_i;\psi) + cV^{1/2}(Q_q(\mathbf{x}_i;\psi))\rfloor$ and $w_q(r_{iq}) = [q\,I(r_{iq} > 0) + (1-q)I(r_{iq} \leqslant 0)]$. The regression coefficients are estimated by the Fisher scoring. As in previous Section, the parameter θ_q is estimated by a robust method:

$$n^{-1}\sum_{i=1}^{n}\left\{\psi_q^2(r_{iq}) - E\left[\psi_q^2\left(\frac{Y_i - Q_q(\mathbf{x}_i;\psi)}{V^{1/2}(Q_q(\mathbf{x}_i;\psi))}\right)\right]\right\} = \mathbf{0}, \tag{12}$$

where $E\left[\psi_q^2\left(\frac{Y_i - Q_q(\mathbf{x}_i;\psi)}{V^{1/2}(Q_q(\mathbf{x}_i;\psi))}\right)\right]$ is a constant that ensures Fisher consistency for the estimation of θ_q and ψ_q can be chosen to be the same as that in (11). The parameter θ_q is estimated by using residuals of Poisson M-quantile model at quantile qth and then, given this estimate, $\boldsymbol{\beta}_q$ is obtained by solving (11). Alternative procedures can be implemented and this is an avenue for future research. Routines `R` for estimation and inference on M-quantile regression models for overdispersed count data data are available from the authors.

A drawback for all quantile-type fitted regression functions is the phenomenon of quantile crossing. This occurs when two or more estimated quantile or M-quantile functions cross or overlap. He (1997) proposed a posteriori restrict quantile regression to avoid the occurrence of crossing and Pratesi *and others* (2009) adapted this procedure to p-splines M-quantile regression. The issue of quantile crossing is addressed by adapting the posteriori technique proposed by He (1997) to obtained NBMQ curves do not cross. This could be a direction for future research.

Hierarchical Bayesian models assume that variability associated with the conditional distribution of Y given x can be explained (at least partially) by spatially structured and unstructured terms (see Section 2). An alternative approach to modelling the variability in this conditional distribution is via NBMQ regression, which does not depend on a hierarchical structure. A key concept in

the application of NBMQ methods to data in disease mapping is identification of a unique 'M-quantile coefficient' associated with each observed datum.

For area i with values y_i and $\mathbf{x}_i$, the area-specific M-quantile coefficient is the value q_i such that $\hat{Q}_q(\mathbf{x}_i;\psi) \approx y_i$. The M-quantile coefficients are estimated by defining a fine grid of values on the interval $(0,1)$ and using the data to fit the NBMQ regression models at each value q on this grid. In the continuous y case the M-quantile coefficient for area i is simply defined as the unique solution q_i to the equation $\hat{Q}_q(\mathbf{x}_i;\psi) = y_i$. However,with count data and $Q_q(\mathbf{x}_i;\psi)$ defined by (10), observed values of y_i can never be part of the strictly positive domain of $Q_q(\mathbf{x}_i;\psi)$ Chambers *and others* (2012). To overcome this problem we suggest the following definition:

$$\hat{Q}_{q_i}(\mathbf{x}_i;\psi) = \begin{cases} k(\mathbf{x}_i) & y_i = 0 \\ y_i & y_i = 1,2,\ldots \end{cases}$$

where $k(\mathbf{x}_i)$ denotes an appropriate strictly positive boundary function for the data set. Note that this cannot be its convex hull, since that will take the value zero where $y_i = 0$. A possibility is $k(\mathbf{x}_i) = \hat{Q}_{q_{min}}(\mathbf{x}_i;\psi)$ where q_{min} denotes the smallest q-value in the grid of q-values used to determine the q_i values of the observed units. However this mean that $q_i = q_{min}$ whenever $y_i = 0$, irrespective of the value of $\mathbf{x}_i$, which seems wrong. One way to tackle this is to follow the same line of argument that Chambers *and others* (2012) used in motivating the definition of q_i for the Bernoulli case. This implies that an observation with value $y_1 = 0$ corresponds to a smaller q-value then another with value $y_2 = 0$ when $\hat{Q}_{0.5}(\mathbf{x}_1;\psi) > \hat{Q}_{0.5}(\mathbf{x}_2;\psi)$. This suggests that we define $k(\mathbf{x}_i) = min\{1-\epsilon, [\hat{Q}_{0.5}(\mathbf{x}_i;\psi)]^{-1}\}$, where $\epsilon > 0$ is a small positive constant. This value can be fixed equal to $-median(\mathbf{x}_i^T\boldsymbol{\beta}_{0.5})$, $i = 1,\ldots,n$, so that half the points with y = 0 have $q > 0.5$ and the remainder have $q \leqslant 0.5$. Then the M-quantile coefficient for area i is q_i, where

$$\hat{Q}_{q_i}(\mathbf{x}_i;\psi) = \begin{cases} min\Big\{1-\epsilon, \frac{1}{t_i exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}}_{0.5})}\Big\} & y_i = 0 \\ y_i & y_i = 1,2,\ldots \end{cases}$$

Note that under (10), solution of the above equation is identical to solution of

$$y_i^\star = \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_{q_i} = \begin{cases} min\Big\{\ln(1-\epsilon)+\ln(t_i), -\mathbf{x}_i^T\hat{\boldsymbol{\beta}}_{0.5}\Big\} - 2\ln(t_i) & y_i = 0 \\ \ln(y_i) - \ln(t_i) & y_i = 1,2,\ldots \end{cases}$$

For a detailed discussion see Chambers *and others* (2012).

The NBMQ modelling approach captures the residual between-area variation by the deviation of the area-specific M-quantile regression coefficient $\boldsymbol{\beta}_{q_i}$ from the 'median' M-quantile coefficient $\boldsymbol{\beta}_{0.5}$. Following Chambers and Tzavidis (2006) this allows to write the NBMQ regression model in a form that mimics the Hierarchical Bayesian Models form, via the identity

$$Q_q(\mathbf{x}_i;\psi) = t_i \exp(\mathbf{x}_i^T\boldsymbol{\beta}_{0.5} + \mathbf{x}_i^T(\boldsymbol{\beta}_{q_i} - \boldsymbol{\beta}_{0.5})). \tag{13}$$

The last term on the right-hand side of (13) can be interpreted as a pseudo-random effect for area i and it allows for capturing the area effects. The M-quantile predictor of the rate in area i is then

$$\hat{Q}_{q_i}(\mathbf{x}_i;\psi) = t_i \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{q_i}). \tag{14}$$

For estimating the MSE of the $\hat{Q}_{q_i}(\mathbf{x}_i;\psi)$ we propose a nonparametric bootstrap-based estimator (hereafter NPB) by constructing a finite artificial population of observed deaths count that imitates the real population, and then from it the bootstrap estimators can be obtained. This procedure is an extension for count data of the method proposed by Chambers *and others* (2012).

Given the finite population of values y_i (a count variable) the steps of the nonparametric bootstrap procedure are summarized as follows:

step 1. Fit model (10) to the initial data and obtain predictors $\hat{Q}_{q_i}(\mathbf{x}_i;\psi)$. For each area compute the pseudo-random effect $\hat{u}_i^{NBMQ} = \bar{\mathbf{x}}_i^T(\hat{\boldsymbol{\beta}}_{q_i} - \hat{\boldsymbol{\beta}}_{0.5})$ and the $\hat{\theta}_{q_i}^{NBMQ}$ at $q = q_i$. It is convenient to re-scale the elements $\hat{\mathbf{u}}^{NBMQ}$ so that they have sample mean exactly equal to zero.

step 2. Sample with replacement from a set $\{1,\ldots,n\}$ in order to construct the vector $\hat{\mathbf{u}}^{NBMQ*} = \{\hat{u}_1^{NBMQ*},\ldots,\hat{u}_n^{NBMQ*}\}^T$ and $\hat{\boldsymbol{\theta}}^{NBMQ*} = \{\hat{\theta}_{q_1}^{NBMQ*},\ldots,\hat{\theta}_{q_n}^{NBMQ*}\}^T$. In particular, $\hat{u}_i^{NBMQ*} = \hat{u}_h^{NBMQ}$ and $\hat{\theta}_{q_i}^{NBMQ*} = \hat{\theta}_{q_h}^{NBMQ}$ where $h = \text{srswr}(\{1,...,n\},1)$.

step 3. Generate a bootstrap data of size n, by generating values of a Negative Binomial distribution with

$$\mu_i^* = t_i \exp\{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{0.5} + \hat{u}_i^{NBMQ*}\},$$

$$\theta_{q_i}^* = \hat{\theta}_{q_i}^{NBMQ*},\ i = 1,\ldots,n,$$

and obtain the bootstrap quantity of interest y_i^*, $i = 1,\ldots,n$.

step 4. Fit model (10) to the bootstrap data and calculate bootstrap predictor $\hat{Q}_{q_i}^*(\mathbf{x}_i;\psi),\ i = 1,\ldots,n$.

step 5. Repeat steps 2-4 B times. In the bth bootstrap replication, let $y_i^{*(b)}$ be the quantity of interest for area i, $\hat{Q}_{q_i}^{*(b)}(\mathbf{x}_i;\psi)$ be the bootstrap Negative Binomial M-quantile predictor.

step 6. A bootstrap estimator of MSE is

$$mse^{NPB}(\hat{Q}_{q_i}(\mathbf{x}_i;\psi)) = B^{-1}\sum_{b=1}^{B}\left(\hat{Q}_{q_i}^{*(b)}(\mathbf{x}_i;\psi) - y_i^{*(b)}\right)^2. \tag{15}$$

5 Real example

Clayton and Kaldor (1987) and many others (e.g. Breslow and Clayton (1993) and Wakefield (2007)) analyzed observed and expected numbers of male lip cancer incidence data collected in the 56 administrative areas of Scotland over the period 1975-1980. The data consist of the observed and expected numbers of cases (based on the age population in each county) and a covariate measuring the proportion of the population engaged in agriculture, fishing, or forestry (AFF). This covariate has been chosen because all three occupations involve outdoor work, exposure to sunlight, the principal known risk factor for lip cancer. Values for the exposure variable AFF are 0 (for 5 areas), 1 (11 areas), 7 (14 areas), 10 (12 areas), 16 (10 areas) and 24 (4 areas). The values are just six, since it was read from a map key, so AFF is a typical measured with error covariate (as suggested also by Wakefield (2007)).

In the present paper, we analyse this data using EB, BYM (without spatially structured random terms) and NBMQ models. As in many other papers, we consider the covariate values divided by ten.

Estimates have been obtained using `R` software. For EB the `eBayes` function on `SpatialEpi` library has been used, while for BYM model the `BRugs` library (an `R` interface to the `OpenBUGS` software) has been used. For NBMQ model we have used an original `R` function, `glm.mq.nb` on `CountMQ` library available from the authors.

Figure 1 shows how for each area we define the belonging quantile using a first estimation M-quantile procedure considering a fine grid from 0.10 to 0.90, by step of 0.05. For a clearer representation, we report in Figure 1 only three quantiles $q = \{0.25, 0.50, 0.75\}$. We consider that each area is ascribed to the estimated quantile of the conditional distributions which is closest to SMR (Y_i/E_i) of the area itself. Geographical location of each area is showed on the right hand of the Figure 1. Then, for each area estimation using NBMQ model for the belonging quantile has been performed.

Figure 2 describes the conditional distribution of number of cases of male lip cancer at different quantiles. In each plot of Figure 2, a dashed line interpolates 17 point estimates (filled dots) of $\boldsymbol{\beta}_{qj}$, $0.010 \leqslant q \leqslant 0.90$, $j = 0, 1$. The effect due to the AFF affects the distribution at the tails. Most notably, the estimate of the parameter 'jumps' by a 50% increase in magnitude within the relatively short interval of quantile points comprised between 0.65 and 0.80.

Figure 3 shows estimates of relative risk for considered models (EB, BYM and NBMQ). Results from different models are quite similar: estimates values lies closed the diagonal. Correlation between estimates obtained from EB and NBMQ is 0.97. Figure 4 describes the box-plot for relative risks estimates using different methods. Negative Binomial M-quantile model seems to have a minor oversmoothing effects than random effects models (behavior criticized among other by Louis (1984), Ghosh (1992) and Green and Richardson (2002)). Figure 5 reports the maps for relative risks estimates confirming the plausibility of suggested NBMQ model.

6 Model-Based Simulation Experiment

We carry out a model-based simulation experiment to compare the performance of the different methods for estimating relative risks in disease mapping. The simulated data are generated from model

$$y_i = t_i \exp\left(-0.35 + 0.72x_i + u_i\right)$$

using expected cases t_i, covariate x_i of the previous example and values of the fitted coefficients $(-0.35, 0.72)$ using EB. Here u_i is drawn from a normal distribution with zero mean and σ^2 equal to 0.15 or 0.25. In the simulation $K = 1,000$ samples for counts data are generated and each sample, $k = 1, \ldots, K$, is perturbed for values of -0.8 on the measure of the covariate on four areas (chosen randomly from the 51 that present a value for the covariate greater than 0.8).

Then, different models are fitted to each sample j for estimating the relative risks for disease mapping: the Negative Binomial linear models under an M-quantile approach (NBMQ), the Empirical Bayesian approach (EB) and the fully Bayesian non spatially structured model (BYM).

The performances of different estimators is evaluated with respect to two basic criteria: the bias (Bias) and the root mean squared error (RMSE). In more details, the Bias is computed, for estimator $\hat{\delta_{ik}}$ for k-th sample of the δ_i parameter in area i as:

$$\text{Bias}_i = \frac{1}{K}\sum_{k=1}^{K}(\hat{\delta}_{ik} - \delta_i),$$

where j indicates the iteration number, and the RMSE is obtained as

$$\text{RMSE}_i = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{\delta}_{ik} - \delta_i)^2}.$$

The mean values of Bias and RMSE over areas are set out in Table 1. The bias results reported in Table 1 confirm our expectations: under both scenarios (σ^2=0.15, 0.25) the EB and the BYM perform better than the NBMQ predictors, whereas from the RMSE results we see that claims in the literature Chambers and Tzavidis (2006) about the superior outlier robustness of M-quantile compared with the EB and BYM certainly hold true. A smearing-type estimator could reduce the bias of the NBMQ predictor. Note, however, that the cost of this bias-adjusted estimator is the inconsistency of (14), reflecting the usual bias-variance trade-off in outlier-robust estimation.

We now examine the performance of the nonparametric MSE estimator (15). Figure 6 shows the area-specific values of RMSE and average-estimated RMSE in case of $\sigma^2 = 0.15,\ 0.25$. Estimator (15) performs well: it track the true RMSE, exhibiting a good stability and showing only a small amount of over-coverage.

7 Negative Binomial M-quantile Geographically Weighted regression

In environmental and epidemiological applications, observations that are spatially close may be more alike than observations that are further apart. One approach for incorporating spatial information in a NBMQ regression model is to assume that the model coefficients themselves vary spatially across the geography of interest. Geographically Weighted Regression (GWR) (Fotheringham *and others* (1997); Fotheringham *and others* (2002); Yu and Wu (2004)) models this spatial variation by using local rather than global parameters in the regression model. The GWR model is a linear model for the conditional expectation of y given $\mathbf{X}$ at location u. That is, a GWR model assumes spatial non-stationarity of the conditional mean of the variable of interest. In this Section we define a spatial extension to NBMQ regression based on Geographically Weighted regression by using the same approach by Salvati *and others* (2012) for M-quantile GWR model. Given n observation at a set of L locations $\{u_l; l = 1, \ldots, L; L \leqslant n\}$, with n_l data values $\{y_{il}, \mathbf{x}_{il}; i = 1, \ldots, n_l\}$ observed at location u_l, the NBMQGWR model can be defined by extending (10) to specify a log-linear model for the M-quantile of order q of the conditional distribution of Y given $\mathbf{X}$ at location u, writing

$$Q_q(\mathbf{X}; \psi, u) = \mathbf{t} \exp(\boldsymbol{\eta}_q(u)), \tag{16}$$

where $\boldsymbol{\eta}_q(u) = \mathbf{X}\boldsymbol{\beta}_q(u)$, the linear predictor, varies with u as well as with q. The parameter $\boldsymbol{\beta}_q(u)$ can be estimated by solving the following estimating equations

$$\Psi(\boldsymbol{\beta}_q(u)) := n^{-1} \sum_{l=1}^{L} w(u_l, u) \sum_{i=1}^{n_l} \boldsymbol{\psi}_q \left(y_i, Q_q(\mathbf{x}_i; \psi, u)\right) = \mathbf{0}, \tag{17}$$

where $\boldsymbol{\psi}_q(y_i, Q_q(\mathbf{x}_i; \psi, u) = \left\{ \psi_q(r_{ilq}) w(\mathbf{x}_{il}) \frac{Q'_q(\mathbf{x}_{il}; \psi, u)}{V^{1/2}(Q_q(\mathbf{x}_{il}; \psi, u))} - a(\boldsymbol{\beta}_q(u)) \right\}$, $w(u_l, u)$ is a spatial weighting function whose value depends on the distance from sample location u_l to u in the sense that sample observations with locations close to u receive more weight than those further away. In this paper we use a Gaussian specification for this weighting function

$$w(u_l, u) = \exp\left\{ - d^2_{u_l, u} / 2b^2 \right\}, \tag{18}$$

where $d_{u_l, u}$ denotes the Euclidean distance between u_l and u and $b > 0$ is the bandwidth. As the distance between u_l and u increases the spatial weight decreases exponentially. The bandwidth b is a measure of how quickly the weighting function decays with increasing distance, and so determines the 'roughness' of the fitted GWR function. A spatial weighting function with a small bandwidth will typically result in a rougher fitted surface than the same function with a large bandwidth. Spatial weights that vary with q can be defined by allowing the bandwidth underpinning these weights to vary with q. Such a q-specific optimal bandwidth b can be obtained by minimising the following function with

respect to b

$$\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left[y_{il}-\hat{Q}_q(\mathbf{x}_{il};\psi,u_{il},b)\right]^2, \tag{19}$$

where $\hat{Q}_q(\mathbf{x}_{il};\psi,u,b)$ is the estimated value of the right hand side of (17) at quantile q and location u_{il}, using bandwidth b when the observation y_{il} is omitted from the model fitting process. However, using this q-specific cross-validation criterion can significantly increase the computational time. For this reason, in this paper we use the optimal bandwidth at $q = 0.5$ for all other values of q for our extension of GWR to NBMQ regression. That is, the spatial weights $w(ul, u)$ in (17) do not depend on q. This is a reasonable first approximation to the q-specific optimal bandwidth that can be computed reasonably quickly, even if this choice could potentially lead to oversmoothing for small or large values of q and hence bias. The parameter $\theta_q(u)$ depends from the location and can be estimated by solving

$$n^{-1}\sum_{l=1}^{L}w(u_l,u)\sum_{i=1}^{n_l}\left\{\psi_q^2(r_{ilq})-E\left[\psi_q^2\left(\frac{Y_{il}-Q_q(\mathbf{x}_{il};\psi,u)}{V^{1/2}(Q_q(\mathbf{x}_{il};\psi,u))}\right)\right]\right\}=\mathbf{0}, \tag{20}$$

where the expectation $E[\cdot]$ in (20) is a constant that ensures Fisher consistency for the estimation of $\theta_q(u)$. The parameter $\theta_q(u)$ varies by locations and quantiles. To reduce the computational time, an alternative approach is to use a value of global $\theta_q(u)$ that varies by quantiles, but it is fixed over space. For example, the θ_q estimated by equation (12). Given the spatial weights and $\theta_q(u)$, the regression coefficients are estimated by the Fisher scoring for each location u.

The M-quantile predictor of the rate in area i is then

$$\hat{Q}_{q_{il}}(\mathbf{x}_{il};\psi,u_{il})=t_{il}\exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}}_{q_{il}}(u_{il})), \tag{21}$$

where fitted regression surface $\hat{Q}_{q_{il}}(\mathbf{x}_{il};\psi,u_{il})=\mathbf{x}_{il}^T\hat{\boldsymbol{\beta}}_q(u_{il})$ defines the fit of the NBMQGWR model for the NB regression M-quantile of order q_{il} of y_{il} given $\mathbf{x}_{il}$ at location u_{il}. Here q_{il} is the M-quantile GWR coefficient for area i with values y_{il} and $\mathbf{x}_{il}$ at location u_{il}. It is calculated as the unique value q_{il} such that $\hat{Q}_{q_{il}}(\mathbf{x}_{il};\psi,u_{il})\approx y_{il}$.

We have considered the Scottish real example dataset. For NBMQGWR we used informations about centroids of each area under the Great Britain National Grid projection system. For BYMspatial model a `BRugs` library (a `R` interface to the `OpenBUGS` software) has been used. For NBMQGWR we have used an original `R` function available from the authors. Figure 7 shows the relative risk maps for both spatially structured models. Figure 8 describes as using NBMQGWR the relative risks for some areas became more similar then obtained from BYMspatial: areas with higher relative risks considering spatial structure move on the diagonal.

8 Conclusion

In this paper, M-quantile models for ecological analysis on disease mapping are introduced and investigated. These models offer a natural way of modeling between area variability in data without imposing prior assumptions about the source of this variability. In particular, with M-quantile models there is no need to explicitly specify the random components of the model; rather, inter-area differences are captured via area-specific M-quantile coefficients. As a consequence, the M-quantile approach reduces the need for parametric assumptions. In addition, estimation and outlier robust inference under these models is straightforward. The proposed approach appears to be suitable for estimating a wide range of parameters and our simulation results show that it is a reasonable alternative to mixed effects models for ecological analysis on disease mapping. A class of robust Negative Binomial models and their extension to M-quantile approach have been proposed. The application to disease mapping and the simulation experiment reveal their applicability. To take the spatial structure of data into account, we consider the possibility to define a Negative Binomial M-quantile Geographically Weighted Regression model. The choice of the model is crucial for the results of disease mapping. Further research seems to be necessary in order to develop tools for model selection of NBMQ based on robust quasi-deviance.

Appendix

We have to evaluate:

$$(i)E\left[\psi_c\left(\frac{Y_i-\mu_i}{V^{1/2}(\mu_i)}\right)\right];\ (ii)E\left[\psi_c\left(\frac{Y_i-\mu_i}{V^{1/2}(\mu_i)}\right)\frac{Y_i-\mu_i}{V(\mu_i)}\right];\ (iii)E\left[\psi_c^2\left(\frac{Y_i-\mu_i}{V^{1/2}(\mu_i)}\right)\right];$$

where Y_i is distributed according to a NEGBIN2 distribution (see Cameron and Trivedi (1998)), that is,

$$P(Y_i=y_i)=\frac{\Gamma(y_i+\theta)}{\Gamma(\theta)\,y_i!}\left(\frac{\mu_i}{\mu_i+\theta}\right)^{y_i}\left(\frac{\theta}{\mu_i+\theta}\right)^{\theta}\ \text{for } y_i=0,1,2,\ldots$$

Here, θ is a positive integer, $\mu_i=E(Y_i)$ and $V(\mu_i)=\mathrm{var}(Y_i)=\mu_i+\frac{\mu_i^2}{\theta}$. Such Y_i can be regarded as the number of failures for having θ successes in a sequence of Bernoulli trials. From now on, to make the notation easier, the index i is suppressed. Accordingly, we write μ instead of μ_i, y instead of y_i, and so on.

First, we evaluate $E\big[Y\,I(Y\in A)\big]$ and $E\big[Y^2\,I(Y\in A)\big]$, where $A=\{a,\ldots,b-1\}$, $0\leqslant a<b$ are integers. Let $A+1=\{a+1,\ldots,b\}$. Then,

$$E\big[Y\,I(Y\in A+1)\big]=E\big[Y\,I(Y\in A)\big]-a\,P(Y=a)+b\,P(Y=b).$$

By the transformation $z = y - 1$, one also obtains

$$\begin{aligned} E\big[\, Y\, I(Y \in A+1)\big] &= \textstyle\sum_{y\in A+1} y\, \frac{\Gamma(y+\theta)}{\Gamma(\theta)\, y!} \left(\frac{\mu}{\mu+\theta}\right)^y \left(\frac{\theta}{\mu+\theta}\right)^\theta \\ &= \frac{\mu}{\mu+\theta}\, \textstyle\sum_{z\in A}(z+\theta)\, \frac{\Gamma(z+\theta)}{\Gamma(\theta)\, z!} \left(\frac{\mu}{\mu+\theta}\right)^z \left(\frac{\theta}{\mu+\theta}\right)^\theta \\ &= \frac{\mu}{\mu+\theta}\, \textstyle\sum_{z\in A}(z+\theta)\, P(Y=z) \\ &= \frac{\mu}{\mu+\theta}\, E\big[\, Y\, I(Y \in A)\big] + \frac{\mu\,\theta}{\mu+\theta}\, P(Y \in A). \end{aligned}$$

Equating such expressions, one finally obtains

$$E\left[Y\, I(Y \in A)\right] = \frac{\mu+\theta}{\theta}\, \left[aP(Y=a) - bP(Y=b)\right] + \mu\, P(Y \in A). \tag{22}$$

We next apply the same argument for calculating $E\left[Y^2\, I(Y \in A)\right]$. Then,

$$E\left[\, Y^2\, I(Y \in A+1)\right] = E\left[Y^2\, I(Y \in A)\right] - a^2\, P(Y=a) + b^2\, P(Y=b)$$

and

$$\begin{aligned} E\left[Y^2\, I(Y \in A+1)\right] &= \textstyle\sum_{y\in A+1} y^2\, \frac{\Gamma(y+\theta)}{\Gamma(\theta)\, y!} \left(\frac{\mu}{\mu+\theta}\right)^y \left(\frac{\theta}{\mu+\theta}\right)^\theta \\ &= \frac{\mu}{\mu+\theta}\, \textstyle\sum_{z\in A}(z+1)(z+\theta)\, \frac{\Gamma(z+\theta)}{\Gamma(\theta)\, z!} \left(\frac{\mu}{\mu+\theta}\right)^z \left(\frac{\theta}{\mu+\theta}\right)^\theta \\ &= \frac{\mu}{\mu+\theta}\, \textstyle\sum_{z\in A}\big(z^2 + (\theta+1)z + \theta\big)\, P(Y=z) \\ &= \frac{\mu}{\mu+\theta}\, E\big[Y^2\, I(Y \in A)\big] + \frac{\mu\,(\theta+1)}{\mu+\theta}\, E\big[\, Y\, I(Y \in A)\big] + \frac{\mu\,\theta}{\mu+\theta}\, P(Y \in A). \end{aligned}$$

Again, equating the above expressions yields

$$\begin{aligned} E\left[Y^2\, I(Y \in A)\right] &= \tfrac{\mu+\theta}{\theta}\, \left[a^2 P(Y=a) - b^2 P(Y=b)\right] + \tfrac{\mu(\theta+1)}{\theta}\, E\left[Y\, I(Y \in A)\right] + \mu\, P(Y \in A) = \\ &= \tfrac{\mu}{\theta}\, \left[\theta + \mu\theta + \mu\right] P(Y \in A) + \tfrac{\mu+\theta}{\theta}\, \left[a^2\, P(Y=a) - b^2\, P(Y=b)\right] + \\ &+ \tfrac{\mu(\mu+\theta)(\theta+1)}{\theta^2}\, \left[aP(Y=a) - bP(Y=b)\right]. \end{aligned} \tag{23}$$

Finally, the previous results, for a particular choice of A allow to evaluate (i)-(ii)-(iii).

Define

$$\psi_c(r) = \begin{cases} r & -c \leqslant r \leqslant c \\ c & r > c \\ -c & r < -c \end{cases}$$

and set $r = \frac{Y-\mu}{V^{1/2}(\mu)}$.

Let $j_1 = \lfloor \mu - c\, V^{1/2}(\mu) \rfloor$ and $j_2 = \lfloor \mu + c\, V^{1/2}(\mu) \rfloor$. Final results may change depending on whether $\mu - c\, V^{1/2}(\mu)$ is integer or noninteger. In what follows, we consider the not integer case; the integer case can be handled similarly.

i)

$$E\left[\psi_c\left(\frac{Y-\mu}{V^{1/2}(\mu)}\right)\right] = -c\, P\left(\frac{Y-\mu}{V^{1/2}(\mu)} < -c\right) + c\, P\left(\frac{Y-\mu}{V^{1/2}(\mu)} > c\right) +$$

$$+ E\left[\frac{Y-\mu}{V^{1/2}(\mu)}\; I(-c \leqslant \frac{Y-\mu}{V^{1/2}(\mu)} \leqslant c)\right] =$$

Because $\frac{Y-\mu}{V^{1/2}(\mu)} > c$ means $Y > \mu + c\,V^{1/2}(\mu)$, as Y should be integer, we have $Y \geqslant \lfloor \mu + c\,V^{1/2}(\mu) \rfloor + 1 = j_2 + 1$. Analogously, $\frac{Y-\mu}{V^{1/2}(\mu)} < -c$ means $Y < \mu - c\,V^{1/2}(\mu)$, which, when $\mu - c\,V^{1/2}(\mu)$ is not integer leads to $Y \leqslant \lfloor \mu - c\,V^{1/2}(\mu) \rfloor = j_1$ (when $\mu - c\,V^{1/2}(\mu)$ is integer to $Y \leqslant j_1 - 1$). Moreover, $-c \leqslant \frac{Y-\mu}{V^{1/2}(\mu)} \leqslant c$ means $\mu - c\,V^{1/2}(\mu) \leqslant Y \leqslant \mu + c\,V^{1/2}(\mu)$ which, when $\mu - c\,V^{1/2}(\mu)$ is not integer is $j_1 + 1 \leqslant Y \leqslant j_2$ (when $\mu - c\,V^{1/2}(\mu)$ is integer is $j_1 \leqslant Y \leqslant j_2$). So, we obtain

$$= -c\,P\left(Y \leqslant j_1\right) + c\,P\left(Y \geqslant j_2 + 1\right) +$$

$$+ \frac{1}{V^{1/2}(\mu)} E\left[Y\, I(j_1 + 1 \leqslant Y \leqslant j_2)\right] - \frac{\mu}{V^{1/2}(\mu)} P(j_1 + 1 \leqslant Y \leqslant j_2).$$

Considering $A = \{j_1 + 1, \ldots, j_2\}$ and also that

$$\frac{\mu + \theta}{\theta}(y+1)P(Y = y+1) = \frac{\mu}{\theta} y P(Y = y) + \mu P(Y = y) \tag{24}$$

we obtain

$$E\left[Y\, I(j_1 + 1 \leqslant Y \leqslant j_2)\right] = \mu P(j_1 \leqslant Y \leqslant j_2 - 1) - \frac{\mu}{\theta} j_2 P(Y = j_2) + \frac{\mu}{\theta} j_1 P(Y = j_1) \tag{25}$$

and finally

$$\begin{aligned} E\left[\psi_c\left(\tfrac{Y-\mu}{V^{1/2}(\mu)}\right)\right] \quad &= -c\,P\left(Y \leqslant j_1\right) + c\,P\left(Y \geqslant j_2 + 1\right) + \\ &+ \tfrac{\mu}{V^{1/2}(\mu)} P(Y = j_1)\left(1 + \tfrac{j_1}{\theta}\right) - \tfrac{\mu}{V^{1/2}(\mu)} P(Y = j_2)\left(1 + \tfrac{j_2}{\theta}\right) \end{aligned}$$

ii)

$$E\left[\psi_c\left(\frac{Y-\mu}{V^{1/2}(\mu)}\right)\frac{Y-\mu}{V(\mu)}\right] = -\frac{c}{V(\mu)} E\left[(Y-\mu)\, I(Y \leqslant j_1)\right] + \frac{c}{V(\mu)} E\left[(Y-\mu)\, I(Y \geqslant j_2 + 1)\right] +$$

$$+ \frac{1}{V^{3/2}(\mu)} E\left[(Y-\mu)^2\, I(j_1 + 1 \leqslant Y \leqslant j_2)\right] =$$

$$= \frac{\mu\, c}{V(\mu)} P(Y \leqslant j_1) + \frac{\mu\, c}{V(\mu)} P(Y \leqslant j_2) + \frac{\mu^2}{V^{3/2}(\mu)} P\left(j_1 + 1 \leqslant Y \leqslant j_2\right) +$$

$$- \frac{c}{V(\mu)} E\left[Y\, I(Y \leqslant j_1)\right] - \frac{c}{V(\mu)} E\left[Y\, I(Y \leqslant j_2)\right] - \frac{2\,\mu}{V^{3/2}(\mu)} E\left[Y\, I(j_1 + 1 \leqslant Y \leqslant j_2)\right] +$$

$$+ \frac{1}{V^{3/2}(\mu)} E\left[Y^2\, I(j_1 + 1 \leqslant Y \leqslant j_2)\right]$$

Considering result (25), $A = \{0, \ldots, j_1\}$ in (22) with (24)

$$E\left[Y\, I(Y \leqslant j_1)\right] = -\frac{\mu}{\theta} j_1\, P(Y = j_1) + \mu\, P(Y \leqslant j_1 - 1), \tag{26}$$

and $A = \{0, \ldots, j_2\}$ in (22) with (24)

$$E\left[Y\, I(Y \leqslant j_2)\right] = -\frac{\mu}{\theta} j_2\, P(Y = j_2) + \mu\, P(Y \leqslant j_2 - 1) \tag{27}$$

and again $A = \{j_1 + 1, \ldots, j_2\}$ in (23) with (24)

$$E\left[Y^2\, I(j_1 + 1 \leqslant Y \leqslant j_2)\right] = \frac{\mu}{\theta} j_1^2 P\left(Y = j_1\right) - \frac{\mu}{\theta} j_2^2 P\left(Y = j_2\right) + \tag{28}$$

$$+\mu P\left(j_1 \leqslant Y \leqslant j_2 - 1\right) + \frac{\mu(\theta + 1)}{\theta} E\left[Y\, I(j_1 \leqslant Y \leqslant j_2 - 1)\right]$$

substituting (25, 26, 27, 28)

$$E\left[\psi_c\left(\frac{Y-\mu}{V^{1/2}(\mu)}\right)\frac{Y-\mu}{V(\mu)}\right] = \frac{\mu\, c}{V(\mu)}\left[P(Y = j_1)\frac{j_1 + \theta}{\theta} + P(Y = j_2)\frac{j_2 + \theta}{\theta}\right] +$$

$$+\frac{\mu}{V^{3/2}(\mu)}\left[P(Y = j_1)\frac{j_1}{\theta}\left(\theta + 1 + j_1\right) - P(Y = j_2)\frac{j_2}{\theta}\left(\theta + 1 + j_2\right) + P(j_1 \leqslant Y \leqslant j_2 - 1)\right] +$$

$$+\frac{\mu^2}{V^{3/2}(\mu)}\left\{P(Y = j_1)\left[\frac{j_1 - j_1\theta - \theta^2}{\theta^2}\right] - P(Y = j_2)\left[\frac{j_2 - j_2\theta - \theta^2}{\theta^2}\right] + \frac{1}{\theta}P(j_1 \leqslant Y \leqslant j_2 - 1)\right\}$$

iii)

$$E\left[\psi_c^2\left(\frac{Y-\mu}{V^{1/2}(\mu)}\right)\right] = c^2\left[P(Y \leqslant j_1) + P(Y \geqslant j_2 + 1)\right] +$$

$$+\frac{1}{V(\mu)} E\left[(Y-\mu)^2\, I(j_1 + 1 \leqslant Y \leqslant j_2)\right] =$$

$$= c^2\left[1 - P(j_1 + 1 \leqslant Y \leqslant j_2)\right] + \frac{\mu^2}{V(\mu)} P(j_1 + 1 \leqslant Y \leqslant j_2) +$$

$$-\frac{2\mu}{V(\mu)} E[Y\, I(j_1 + 1 \leqslant Y \leqslant j_2)] + \frac{1}{V(\mu)} E[Y^2\, I(j_1 + 1 \leqslant Y \leqslant j_2)] =$$

Substituting the values of the expected values (25, 28) we arrive to

$$E\left[\psi_c^2\left(\frac{Y-\mu}{V^{1/2}(\mu)}\right)\right] = c^2\ \left[1 - P(j_1 + 1 \leqslant Y \leqslant j_2)\right] +$$

$$+\frac{\mu}{V(\mu)}\left[P(Y = j_1)\ \frac{j_1}{\theta}\left(\theta + 1 + j_1\right) - P(Y = j_2)\ \frac{j_2}{\theta}\left(\theta + 1 + j_2\right) + P(j_1 \leqslant Y \leqslant j_2 - 1)\right] +$$

$$+\frac{\mu^2}{V(\mu)}\left\{P(Y = j_1)\left[\frac{j_1 - j_1\theta - \theta^2}{\theta^2}\right] - P(Y = j_2)\left[\frac{j_2 - j_2\theta - \theta^2}{\theta^2}\right] + \frac{1}{\theta}P(j_1 \leqslant Y \leqslant j_2 - 1)\right\}$$

References

Assuncao, R.M. (2003). Space varying coefficient models for small area data. *Environmetrics* **14**, 453–473.

Besag, J., York, J.C. and Mollié, A. (1991). Bayesian image restoration, with application in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika* **75**, 761–771.

Breslow, N.E. (1984). Extra-poisson variation in log-linear models. *Applied Statistics* **33**, 38–44.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Society* **88**, 9–25.

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge University Press.

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association* **96**, 1022–1030.

Chambers, R., Salvati, N. and Tzavidis, N. (2012). M-quantile regression models for binary data in small area estimation. *Submitted: available from the authors upon request*.

Chambers, R. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika* **93**, 255–268.

Clayton, D. and Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671––681.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - an empirical bayes approach. *Journal of the American Statistical Association* **68**, 117–130.

Fotheringham, A.S., Brundson, C. and Charlton, M. (1997). Two techniques for exploring non-stationarity in geographical data. *Journal of Geographical Systems* **4**, 59–82.

Fotheringham, A.S., Brundson, C. and Charlton, M. (2002). *Geographically Weighted Regression*. West Sussex: John Wiley & Sons.

Ghosh, M. (1992). Constrained bayes estimation with applications. *Journal of the American Statistical Association* **87**, 533–540.

Green, P.J. and Richardson, S. (2002). Hidden markov models and disease mapping. *Journal of the Statistical Association* **97**, 1–16.

He, X. (1997). Quantile curves without crossing. *American Statisticians* **51**, 186–192.

Huber, P.J. (1981). *Robust Statistics*. London: Wiley.

Kensall, J.E. and Wakefield, J.C. (1999). *Discussion of "Bayesian Models for Spatially Correlated Disease and Exposure Data"*, Bayesian Statistics 6 edition. Ney York Oxford University Press.

Kokic, P., Chambers, R., Breckling, J. and Beare, S. (1997). A measure of production performance. *Journal of Business and Economic Statistics* **10**, 419–435.

Lawless, J.F. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics* **15**, 209–225.

Louis, T.A. (1984). Estimating a population of parameter values using empirical bayes methods. *Journal of the American Statistical Association* **79**, 393–398.

MacNab, Y.C. (2009). Bayesian multivariate disease mapping and ecological regression with errors in covariates: Bayesian estimation of dalys and 'preventable' dalys. *Statistics in Medicine* **28**, 1369–1385.

MacNab, Y.C. (2010). On bayesian shared component disease mapping and ecological regression with errors in covariates. *Statistics in Medicine* **29**, 1239–1249.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, second edition edition. London – New York: Chapman and Hall.

Pratesi, M., Ranalli, M.G. and Salvati, N. (2009). Semiparametric m-quantile regression using penalized splines. *Journal of Nonparametric Statistics* **21**, 287–304.

Rao, J.N.K. (2003). *Small Area estimation*. Wiley Series in Survey methodology.

Salvati, N., Tzavidis, N., Pratesi, M. and Chambers, R. (2012). Small area estimation via m-quantile geographically weighted regression. *Test* **21**, 1–28.

Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics* **8**, 158–183.

Yu, D.L. and Wu, C. (2004). Understanding population segregation from landsat etm+imagery: a geographically weighted regression approach. *GISience and Remote Sensing* **41**, 145–164.

Table 1: Model-based simulation results: performances of predictors of relative risks for disease mapping

	σ^2=0.15		σ^2=0.25	
	Bias	RMSE	Bias	RMSE
EB	0.001	0.520	-0.002	0.769
BYM	0.002	0.563	-0.001	0.819
NBMQ	-0.036	0.417	-0.083	0.514

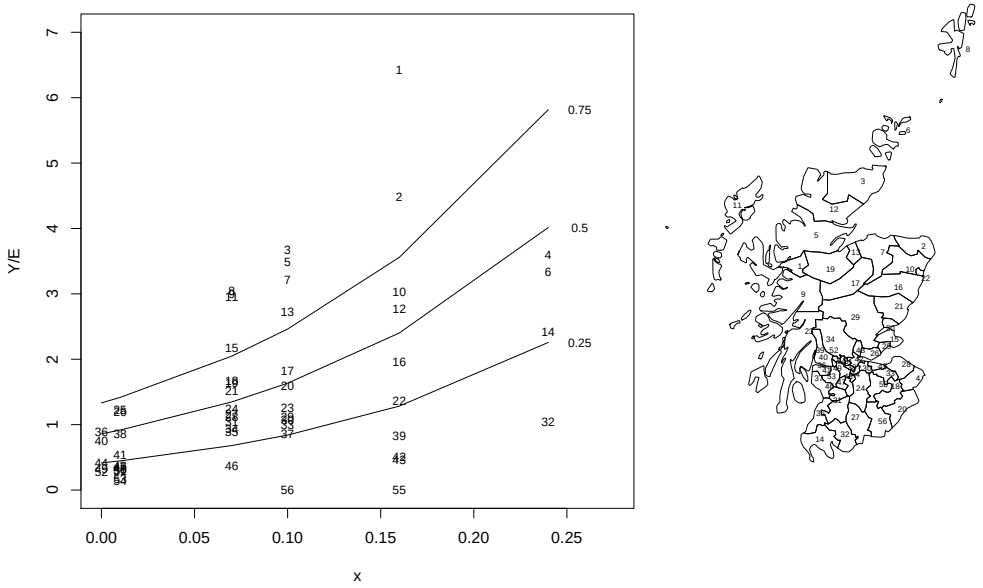

Figure 1: Observed data and predicted values of quantiles $q = \{0.25, 0.50, 0.75\}$ from NBMQ model

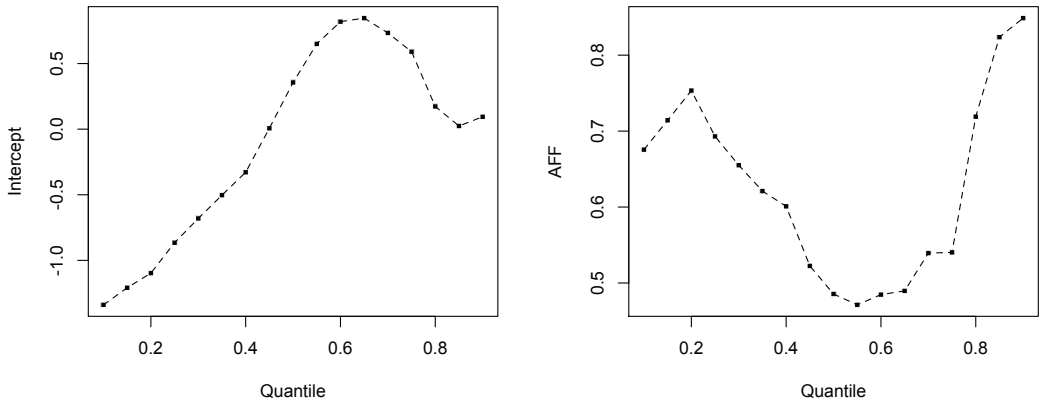

Figure 2: Conditional NBMQ quantiles $q = \{0.10, 0.15, \ldots, 0.50, \ldots, 0.85, 0.90\}$ estimated by model (10)

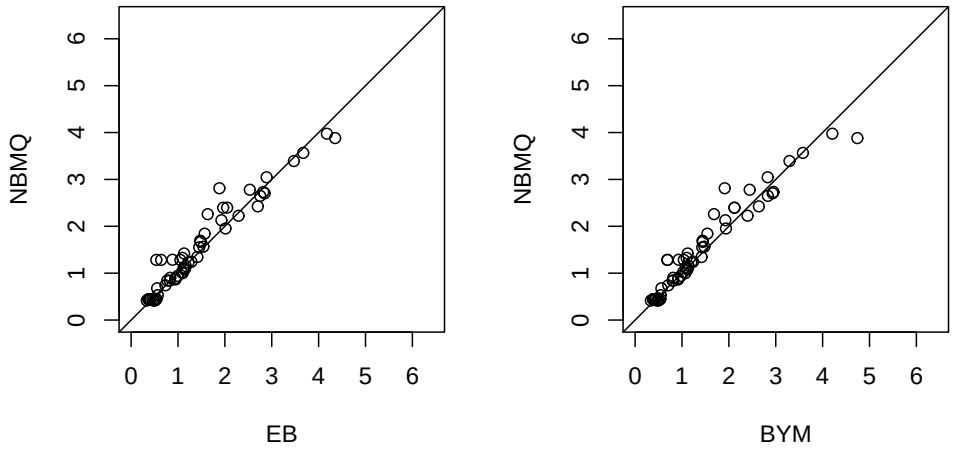

Figure 3: Relative risks estimates using different models: EB, BYM and NBMQ

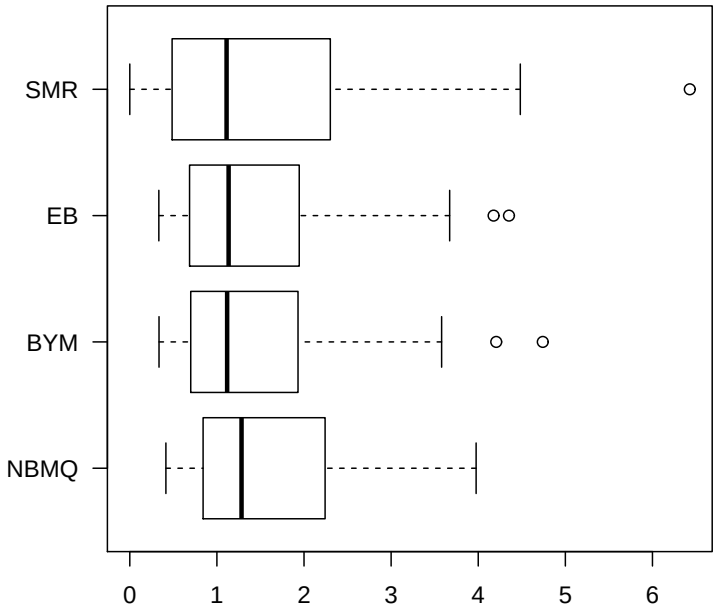

Figure 4: Relative risks estimates using different models: SMR, EB, BYM and NBMQ

Figure 5: Relative risks estimates using models: SMR, EB, BYM and NBMQ

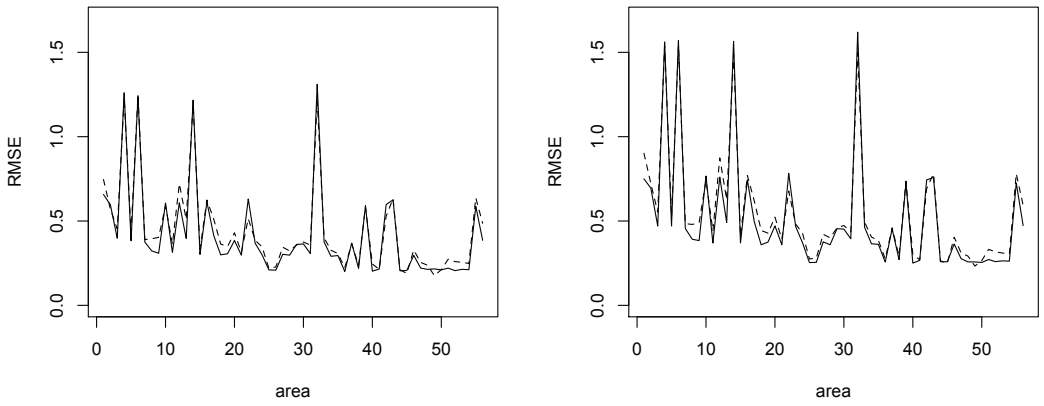

Figure 6: Area-specific values of RMSE (solid line) and average-estimated RMSE (dashed line) under σ^2 equal to 0.15 (left) and 0.25 (right)

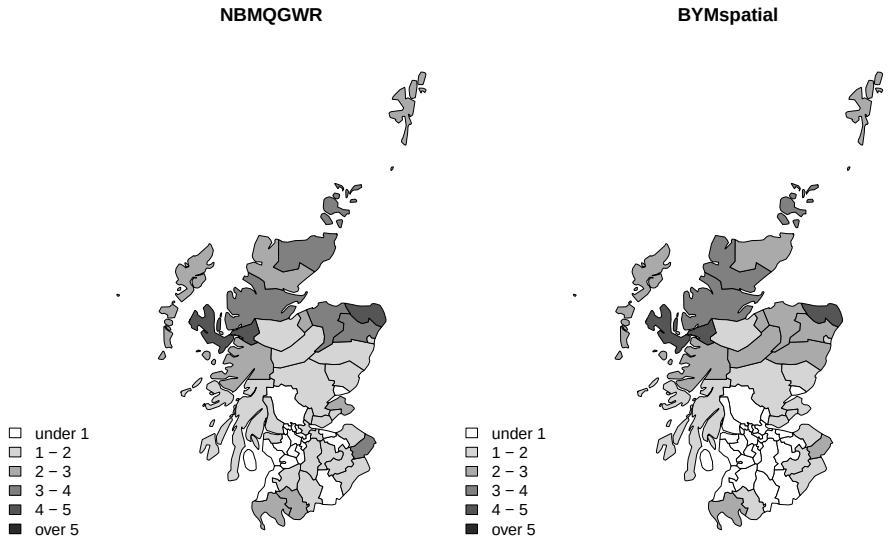

Figure 7: Relative risks estimates using models: NBMQGWR and BYMspatial

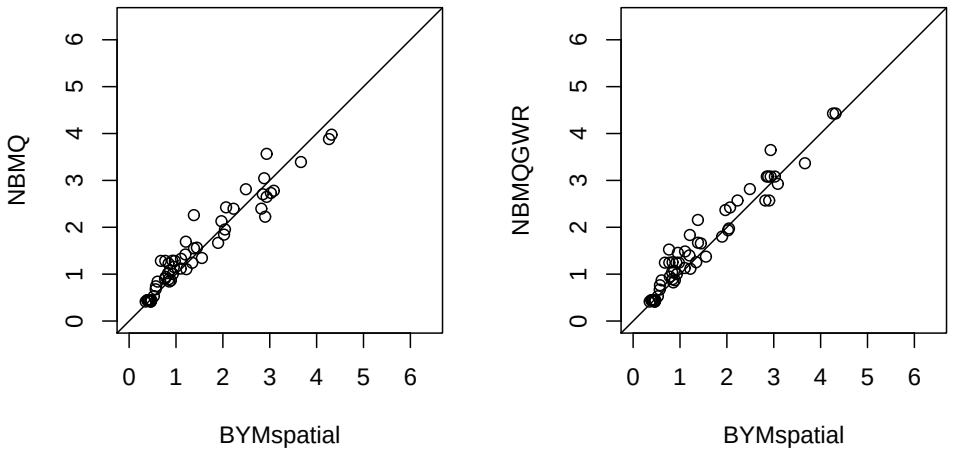

Figure 8: Relative risks estimates using different models: NBMQ, NBMQGWR and BYMspatial