Poisson M-Quantile Geographically Weighted Regression on Disease mapping

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

Emanuela Dreassi
Dipartimento di Statistica "G. Parenti", Università degli Studi di Firenze,
Viale Morgagni, 59 - 50134 Florence, Italy email: dreassi@ds.unifi.it

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

Abstract: A new approach to ecological analysis on disease mapping is introduced: a semi-parametric approach based on M-quantile models. We define a Poisson M-Quantile spatially structured model. The proposed approach is easily made robust against outlying data values for covariates. Robust ecological disease mapping is desirable since covariates at area level usually present measure-type error. We consider a spatial structure in the model in order to extend the M-quantile approach to account for spatial correlation between areas using Geographically Weighted Regression (GWR). Differences between M-quantile and usual random effects models are discussed and the alternative approaches are compared using the Scottish Lip cancer example.

Keywords: disease mapping, ecological analysis, M-quantile regression, Robust models, spatial correlation, Poisson regression, geographically weighted regression

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 aggregate data are often relatively easy to obtain from government sources. More difficult is to obtain the measures, 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 the analysis of the association between risk factors and disease. In ecological analysis related to disease mapping, data usually exhibit overdispersion. The latter is usually considered in the model by way of random effects introduced on the model. Clayton and Kaldor (1987) proposed the use of a Poisson-

gamma model for relative risks using an Empirical Bayesian approach (referred to as EB below). This model was generalized by Besag *et al.* (1991) into a fully Bayesian setting using a Hierarchical Bayesian model with a spatial structure (referred to as BYM below). So, ecological disease mapping typically rely on regression models that use both covariates and random effects to explain variation between areas. These models depend on strong distributional assumptions and require a formal specification of the random part of the model. Moreover, they do not easily allow for outlier-robust inference due to covariates at areal level that could be measure-type error prone.

In this paper, we describe a new approach to ecological disease mapping: Poisson M-Quantile regression (referred to as PMQ below). Raughly speaking, the idea is to model quantiles like parameters of the conditional distribution of the target variable given the covariates. Unlike usual random effects models, M-quantile models do not depend on strong distributional assumptions and are robust to the presence of outliers due to measure-type error on covariates. We introduce easily a spatial structure extending the M-quantile approach to account for such spatial correlation between areas by way of appropriate weights at the estimation step (see Salvati *et al.*, 2011). The used approach to incorporate such spatial information is Geographically Weighted Regression: the relationship between the outcome variable and the covariates is characterised by local rather than global parameters, where local is defined spatially. Differences between Poisson M-quantile and traditional random effects models are discussed and compared using the Scottish Lip cancer example.

2 Poisson M-Quantile regression

We define an extension of linear M-quantile regression to count data. M-quantile regression (Breckling and Chambers, 1988) is a "quantile-like" generalization of regression based on the influence function (M-regression). The M-quantile of order q, $q \in (0,1)$, of a random variable Y with continuous distribution function $F(\cdot)$ is the value Q_q that satisfies

$$E\left[\psi_q\left(\frac{Y-Q_q}{\sigma_q}\right)\right]=0$$

where σ_q is a suitable measure of the scale of the random variable $Y-Q_q$, $\psi_q(\epsilon)=2\psi(\epsilon)\left[qI(\epsilon>0)+(1-q)I(\epsilon\leq 0)\right]$ and ψ is an appropriately chosen influence function: the Huber "small c" second proposal specification with $c=1.345$, $\psi(\epsilon)=\epsilon I(-c\leq\epsilon<c)+c\ sgn(\epsilon)I(|\epsilon|>c)$.

Breckling and Chambers (1988) define a linear M-quantile regression model as one where the M-quantile $Q_q(X;\Psi)$ of the conditional distribution of Y given the matrix of p auxiliary variables X corresponding to an influence function ψ satisfies

$$Q_q(X;\psi)=X\beta_{q\psi}$$

There is no agreed definition of an M-quantile regression function when Y is rates parameterized Poisson. The most appealing, of course, is using a log-linear specification

$$Q_q(X;\psi) = t\exp(\gamma_{q\psi})$$

where $\gamma_{q\psi} = X\beta_{q\psi}$ is the linear predictor and t the offset term (expected cases of death). Cantoni and Ronchetti (2001) obtained a robust version of the estimating equations for generalized linear models. We consider the extensions of this to the M-quantile geographically weighted regression case (referred to as PMQGWR below) following Salvati *et al.* (2011).

3 Scottish Lip cancer Example

Clayton and Kaldor (1987) and many others (i.e. Wakefield, 2007) analyzed observed and expected numbers of lip cancer cases in the 56 administrative areas of Scotland. Data were available on the percentage of the work force in each county employed in agriculture, fishing or forestry. This covariate have been chosen because all three occupations involve outdoor work, exposure to sunlight, the principal known risk factor for lip cancer. In the present paper, we analyse this data using EB, BYM using a convolution prior (exchangeable and spatially structured random terms), PMQ and PMQGWR models. Figure 1 shows estimates of relative risk for considered models. Results are similar. Poisson M-quantile models, seems smoother less than random effects models. For PMQGWR sensitivity analysis to bandwidth choice has to be considered.

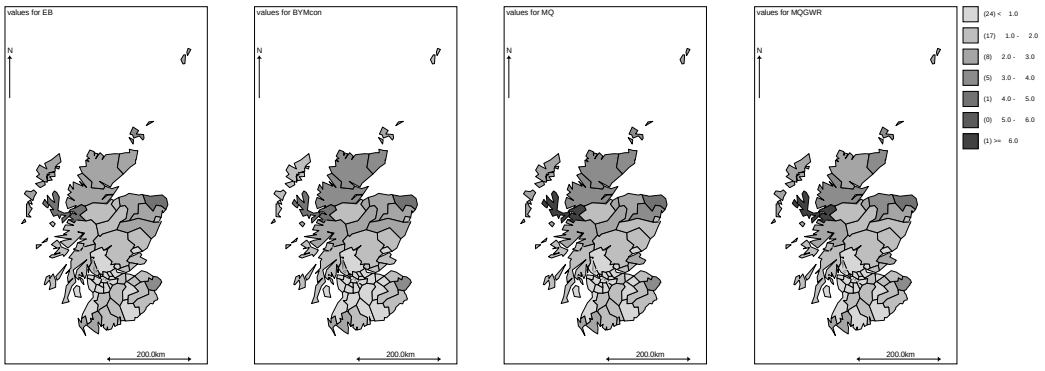

Figure 1: Relative risks estimates using different models: EB, BYM gaussian convolution, PMQ and PMQGWR

4 Conclusion

In this paper, M-quantile models for ecological analysis on disease mapping are introduced and investigated. In particular, we specify an M-quantile GWR model that is a local model for the M-quantiles of the conditional distribution of the outcome variable given the covariates. This model is then used to define a bias-robust predictor of the small area characteristic of interest that also accounts for spatial association in the data. These models offer a natural way of modeling between area association and variability 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.

References

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

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

Cantoni E., Ronchetti E. (2001) Robust Inference for Generalized Linear Models, *Journal of the American Statistical Association*, 96, 1022-1030.

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

Lawson A.B., Browne W.J., Vidal Rodeiro C.L. (2003) *Disease Mapping with WinBUGS and MLwiN*, Wiley, London.

Salvati N., Tzavidis N., Pratesi M., Chambers R. (2011) Small area estimation via M-quantile Geographically Weighted Regression, *Test*, forthcoming (DOI 10.1007/s11749-010-0231-1).

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