2013

Poisson M-quantile regression for small area estimation

Nikos Tzavidis
University of Southampton

Maria Giovanna Rannalli
University of Perugia

Nicola Salvati
University of Pisa

Emanuela Dreassi
University of Florence

Ray Chambers
University of Wollongong, ray@uow.edu.au

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

14-13

Poisson M-quantile Regression for Small Area Estimation

Nikos Tzavidis, M. Giovanna Ranalli , Nicola Salvati, Emanuela Dreassi and Ray Chambers

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Poisson M-quantile Regression for Small Area Estimation

Nikos Tzavidis* M. Giovanna Ranalli† Nicola Salvati‡
Emanuela Dreassi§ Ray Chambers¶

Abstract

A new approach to model-based small area estimation for count outcomes is proposed and used for estimating the average number of visits to physicians for Health Districts in Central Italy. The proposed small area predictor is based on defining a Poisson M-quantile model by extending the ideas in Cantoni & Ronchetti (2001) and Chambers & Tzavidis (2006). This predictor can be viewed as a semi-parametric outlier robust alternative to the more commonly used plug-in Empirical Best Predictor that is based on a Poisson generalised linear mixed model with Gaussian random effects. Results from the real data application and from a simulation experiment confirm that the proposed small area predictor has good robustness properties and can be more efficient than alternative small area predictors.

Keywords: Count data; generalized linear models; health survey; non-normal outcomes; nonparametric bootstrap; robust inference.

1 Introduction

The Health Conditions and Appeal to Medicare Survey (HCAMS) is a national, multistage, personal interview sample survey conducted periodically in Italy by

*Southampton Statistical Sciences Research Institute, University of Southampton, Southampton SO17 1BJ, UK, n.tzavids@soton.ac.uk

†Dipartimento di Economia, Finanza e Statistica, Universtà degli Studi di Perugia, Via Pascoli - I 06123 Perugia, Italy giovanna@stat.unipg.it

‡Dipartimento di Economia e Management, Università di Pisa, Via Ridolfi, 10 - I 56124 Pisa, Italy, salvati@ec.unipi.it

§Dipartimento di Statistica, Informatica, Applicazioni 'G. Parenti', Università degli Studi di Firenze, Viale Morgagni, 59 - I 50134 Firenze, Italy, dreassi@disia.unifi.it

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

the National Institute of Statistics. The 2012-13 edition is currently running; previous editions have been conducted in the periods 1999-2000 and 2004-05. It provides information about the health condition and health care use of the non-institutionalized population of Italy. The questionnaire comprises items on basic health condition (like perceived health status, MBI and dietary habits) that are also surveyed annually by the Multipurpose Everyday Life Survey. In addition, it covers special health topics on chronic and acute diseases, visits to physicians and general practitioners.

HCAMS is a multistage survey in which municipalities are primary sampling units (PSUs), while households are secondary sampling units (SSUs). The 1999-2000 edition has about 1,449 PSUs (out of 8,102) and 52,332 households with approximately 120,000 individuals. HCAMS is designed to provide reliable (direct and design based) estimates at the Administrative Region (NUTS2) level, but there is also a need for estimates for smaller subpopulations or geographical areas. This is true in general for National Surveys, but particularly relevant for surveys with health related information, since in Italy health is mainly managed locally at the level of NUTS2. In particular, policies are endorsed by Administrative Regions by allocating resources and funds to Health Districts (HDs) that are in charge for local implementation. HDs are defined by groups of contiguous municipalities and are not planned domains in the HCAMS. A good number of HDs have very low sample size and represent, therefore, *small areas* of interest.

The increasing demand from the administrators and policy planners for reliable estimates of various parameters at small area level has led to the development of a number of efficient model-based small area estimation (SAE) methods (see Rao, 2003, for a review of such methods). For example, the empirical best linear unbiased predictor (EBLUP) based on a linear mixed model (LMM) is often recommended when the target of inference is the small area average of a continuous response variable (Battese et al., 1988). However, using a LMM to characterise differences between small areas requires strong distributional assumptions. Robust SAE inference under the LMM has recently attracted some interest (Sinha & Rao, 2009; Chambers et al., 2013). An alternative approach to small area estimation that automatically allows for robust inference is to use M-quantile models (Breckling & Chambers, 1988) to characterise these differences (Chambers & Tzavidis, 2006).

Most of the variables in the HCAMS are binary or take the form of a count and are therefore not suited to standard SAE methods based on LMMs. Working within a frequentist paradigm, one can follow Jiang & Lahiri (2001) who propose an empirical best predictor (EBP) for a binary response, or Jiang (2003) who extends these results to generalized linear mixed models (GLMMs). Nevertheless, use of EBP can be computationally challenging (Molina & Rao, 2010). Despite their attractive properties as far as modelling non-normal outcomes is concerned,

the use of GLMMs requires numerical approximations. In particular, the likelihood function defined by a GLMM can involve high-dimensional integrals which cannot be evaluated analytically (see Mc Culloch, 1994, 1997; Song et al., 2005). In such cases numerical approximations can be used, as for example in the `R` function `glmer` in the package `lme4`. Alternatively, estimation of the model parameters can be obtained by using an iterative procedure that combines Maximum Penalized Quasi-Likelihood (MPQL) and REML estimation (Saei & Chambers, 2003). For all these reasons, alternative approaches to modelling discrete outcomes should be considered. Furthermore, estimates of GLMM parameters can be very sensitive to outliers or departures from underlying distributional assumptions. Large deviations from the expected response as well as outlying points in the space of the explanatory variables are known to have a large influence on classical maximum likelihood inference based on generalized linear models (GLMs). Following a Bayesian paradigm, Maiti (2001) describes a Hierarchical Bayes approach to fitting a GLMM based on an outlier-robust normal mixture prior for the random effects and uses this model for SAE. Sinha (2004) proposes robust estimation of the fixed effects and the variance components of a GLMM, using a Metropolis algorithm to approximate the posterior distribution of the random effects.

Let us introduce briefly the notation for small area estimation and GLMMs. Let U denote a finite population of size N which can be partitioned into D domains or small areas, with U_d denoting population on small area d, $d = 1, ..., D$. The small area population sizes N_d, for $d = 1, ..., D$ are assumed known. Let y_{dj} be the value of the outcome of interest, for the purposes of this paper a discrete or a categorical variable, for unit j in area d, and let $\mathbf{x}_{dj}$ denote a $p \times 1$ vector of unit level covariates (including an intercept). It is assumed that the values of $\mathbf{x}_{dj}$ are known for all units in the population, as are the values $\mathbf{z}_d$ of a $q \times 1$ vector of area level covariates. We will see that the first requirement can be relaxed to some extent when there are no continuous variables among the $\boldsymbol{x}$'s. In the presence of categorical covariates, an equivalent alternative representation of the assumed data structure is in the form of a cross-tabulation. The aim is to use the sample values of y_{dj} and the population values of $\mathbf{x}_{dj}$ and $\mathbf{z}_d$ to estimate a proportion or a count of a characteristic in the small area $d = 1, ..., D$.

For discrete outcomes model-based small area estimation conventionally employs a GLMM for $\mu_{dj} = E[y_{dj}|\mathbf{u}_d]$ of the form

$$g(\mu_{dj}) = \eta_{dj} = \mathbf{x}_{dj}^T\boldsymbol{\beta} + \mathbf{z}_d^T\mathbf{u}_d, \tag{1}$$

where g is a link function. When y_{dj} is a count outcome the logarithmic link function is commonly used and the individual y_{dj} values in area d are assumed to be

independent Poisson random variables with

$$\mu_{dj} = E[y_{dj}|\mathbf{u}_d] = \exp\{\eta_{dj}\} \tag{2}$$

and $\text{Var}[y_{dj}|\mathbf{u}_d] = \mu_{dj}$. The q-dimensional vector $\mathbf{u}_d$ is generally assumed to be independently distributed between areas according to a normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_u$. $\mathbf{\Sigma}_u$ depends on parameters $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)$, which are referred to as the variance components and $\boldsymbol{\beta}$ in (1) is the vector of fixed effects. If the target of inference is the small area d mean, $\bar{y}_d = N_d^{-1} \sum_{j \in U_d} y_{dj}$ and the Poisson-GLMM (1) is assumed, the approximation to the minimum mean squared error predictor of $\bar{y}_d$ is $N_d^{-1}[\sum_{j \in s_d} y_{dj} + \sum_{j \in r_d} \mu_{dj}]$. Since μ_{dj} depends on $\boldsymbol{\beta}$ and $\mathbf{u}_d$, a further stage of approximation is required, where unknown parameters are replaced by suitable estimates. This leads to the plug-in version of the EBP (hereafter EBPP) for the area d proportion $\bar{y}_d$ under (2),

$$\hat{\bar{y}}_d^{\text{EBPP}} = N_d^{-1}\Big\{ \sum_{j \in s_d} y_{dj} + \sum_{j \in r_d} \hat{\mu}_{dj} \Big\}, \tag{3}$$

where $\hat{\mu}_{dj} = \exp\{\hat{\eta}_{dj}\}$, $\hat{\eta}_{dj} = \mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_d^T \hat{\mathbf{u}}_d$, $\hat{\boldsymbol{\beta}}$ is the vector of the estimated fixed effects and $\hat{\mathbf{u}}_d$ denotes the vector of the predicted area-specific random effects (see Rao, 2003; Saei & Chambers, 2003; Jiang & Lahiri, 2006; González-Manteiga et al., 2007). In (3) s_d and r_d denote the set of sampled (of size n_d) and non-sampled (of size $N_d - n_d$) units in small area d, respectively.

In this paper, we focus on estimates of the mean number of visits to physicians within the past four weeks among people aged 65 or more for 60 HDs comprised in the three Administrative Regions of Liguria, Toscana and Umbria. These are neighboring Regions placed in the center part of Italy and share a common concern on the quality of services for elderly. Aging of the population is a general concern in Italy given that it shows the largest proportion of people aged 65 or more in Europe (20.3% in 2011, latest available figure). Liguria, Toscana and Umbria are three of the four 'oldest' regions in Italy with proportions of 26.7%, 23.3% and 23.1%, respectively.

Malec et al. (1997) also consider the estimation of quantities related to the number of visits to physicians from the American National Health Interview Survey. They focus on the proportion of population with at least one visit in the past twelve months and use a Hierarchical Bayesian approach. A logistic model to relate the individual's probability of a doctor visit to his/her characteristics is used and then small area parameters are modeled with respect to area specific covariates. In this case, we are interested in the *number* of visits to physicians and, therefore, we need to properly model this count variable. In addition, the distribution of the variable of interest shows some unduly large values that require a robust procedure to properly account for them in the estimation process. It is in

fact very important that final estimates are not overly influenced by them to provide a reliable comparison among small areas.

For all these reasons, in this paper we present a new approach to SAE for count outcomes based on M-quantile modeling. These models do not depend on strong distributional assumptions nor on a predefined hierarchical structure, and outlier robust inference is automatically performed when these models are fitted. Following Chambers et al. (2012b) and Chambers et al. (2012a) we extend the existing M-quantile approach for continuous data to the case where the response is a count. As with M-quantile modeling of a continuous response (Chambers & Tzavidis, 2006) random effects are avoided and between area variation in the response is characterised by variation in area-specific values of quantile-like coefficients. In Section 2 we motivate the use of M-quantile models for the estimation of mean number of visits to the physicians with some explorative analysis on the data set. In Section 3, after reviewing M-quantile small area estimation for a continuous response, we show how the approach for robust inference for GLMs proposed by Cantoni & Ronchetti (2001) can be extended for fitting an M-quantile GLM. Approaches for defining the M-quantile coefficients, which play the role of pseudo-random effects in this framework, are discussed in Section 4, alongside the definition of small area predictors and corresponding Mean Squared Error (MSE) estimators. In Section 5 we report the results from the application of the proposed methodology for deriving estimates of the number of visits in primary health care outlets for HDs in Italy. Results from model-based and design-based simulation studies aimed at empirically assessing the performance of the proposed small area predictors are presented in Section 6. Section 7 concludes the paper with some remarks and possible future researches.

2 The estimation of mean number of visits to a doctor from HCAMS: empirical challenges

The survey design of HCAMS uses a complex sampling scheme and, in particular, it is as follows. Within a given Province (LAU1), municipalities are classified as Self-Representing Areas (SRAs) - consisting of the larger municipalities - and Non Self-Representing Areas (NSRAs) - consisting of the smaller ones. In SRAs each municipality is a single stratum and households are selected by means of systematic sampling. In NSRAs the sample is based on a stratified two stage sample design. Municipalities are PSUs, while households are SSUs. PSUs are divided into strata of approximately the same dimension in terms of population. One PSU is drawn from each stratum with probability proportional to the PSU population size. The SSUs are selected from population registers held by municipalities by

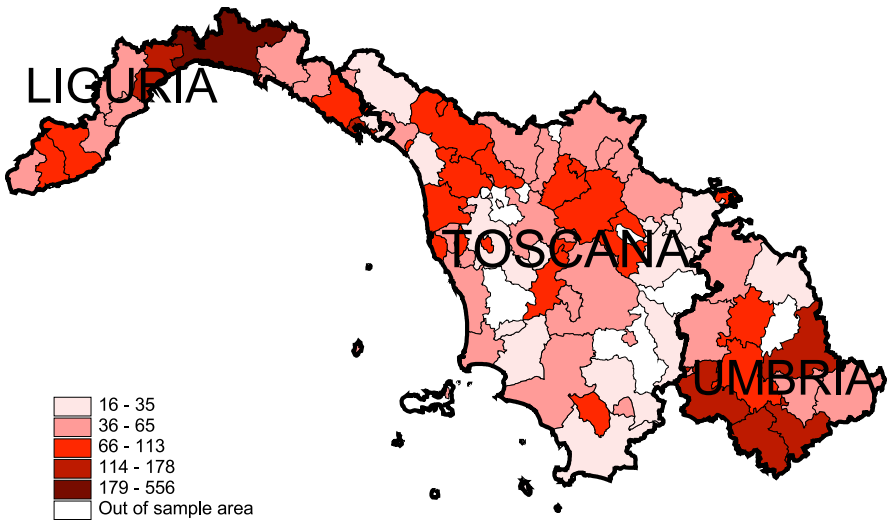

Figure 1: Sample size in each Health Districts of Liguria, Toscana and Umbria in 2000.

means of systematic sampling in each PSU. All members of each sample household, in both SRAs and NSRAs, are interviewed.

The data that we consider in this paper are from the 1999-2000 edition of HCAMS. We are interested in producing estimates of the average number of visits to physicians within the past four weeks for elderly (aged 65 or more) in the 60 HDs of Toscana, Liguria and Umbria. Recall that HDs are groups of neighboring municipalities (PSUs), while elderly is a domain that cuts across PSUs. The total sample size for the three Regions is $n = 4{,}021$. Figure 1 provides a map of the three regions of interest; HDs are color coded according to the sample size. Note that 5 HDs in Toscana and 1 in Umbria are out of sample areas, i.e. they have no sampled units.

The application of small area unit level methodologies requires individual level covariate information for all units in the population for each area. One possible source of these data is the Population Census run in Italy in 2001. Among variables available from the census, we concentrate on those that are also available every year from population registers held at a municipality level. This is the set of auxiliary variables that are available also for the other editions of the survey and can eventually be employed to obtain estimates and have comparisons over time. From administrative registers, we have the distribution for each municipality by age, gender and marital status. We collapse age in 5-year classes (65-69, 70-74, 75-79, 80-84, 85 or more) and also consider an overall Administrative Region effect.

Table 1 reports the results of a procedure of analysis of deviance from fitting

Table 1: Analysis of deviance table from fitting Poisson-Normal mixed models to the whole dataset (in italics the models with nonsignificant improvements in the fit).

Covariates	Resid. df	df	Deviance	p-value
Null	4019			
age	4015	4	59.590	1.5e-11
age, gender	4014	1	12.816	0.0003
age, gender, marital status	*4011*	*3*	*2.277*	*0.5170*
age, gender, region	4012	2	4.196	0.1227
age, gender, region, age $\times$ gender	*4008*	*4*	*1.165*	*0.8838*
age, gender, region, region $\times$ gender	*4010*	*2*	*0.746*	*0.6888*
age, gender, region, region $\times$ age	*4004*	*8*	*8.962*	*0.3455*

a Poisson-Normal mixed effects model on the sample data, in which a random intercept is fitted for each HD. We can note that age class, gender and region are significant. On the other hand, marital status and interactions between pairs of the three significant variables are not significant. A likelihood ratio test for the significance of the variance of the distribution of the small area random effects has been conducted as well. Given that we are testing whether the parameter of interest is zero, i.e. a value on the boundary of its parameter space, it can be shown that minus twice the ratio of the two log-likelihoods has approximate density function equal to a 50:50 mixture between a χ^2_0 and a χ^2_1 distribution. Using this approximation instead of the usual χ^2_1 essentially leads to p-values being halved. In our case the value of the test statistic is 33.695, with a p-value $3.22e-09$, that provides evidence of a significant area effect.

An important property of the Poisson regression model is that it allows to analyze individual or grouped data with equivalent results due to the fact that the sum of independent Poisson random variables is also Poisson. This is useful when we have groups of individuals with identical covariate values as it occurs in this case study. For this reason, in the sequel we will fit models to grouped data in which groups are defined by cross classifying gender by age class by HDs. The response variable is the total number of visits to physicians for all individuals in each group. The overall sample size from each group is considered as an offset.

Figure 2 reports two plots of Pearson residuals from the Poisson-Normal mixed effects model with covariates given by age class, gender and Region. The histogram clearly shows that the distribution of the residuals is positively skewed and has some quite large values. This is confirmed by the second plot, representing the distribution of the residuals by HD: some HDs contain many positive residuals, whereas some HDs have negative residuals. Finally, Figure 3 plots the raw residu-

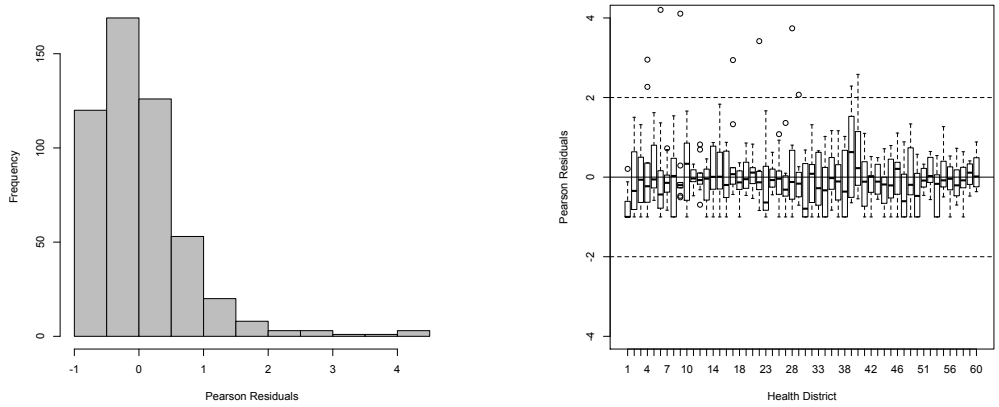

Figure 2: Model fit diagnostics for a Poisson-Normal mixed model fit to the data: histogram of Pearson residuals (left) and box-plots of Pearson residuals by Health District (right).

als against the predicted values for the number of visits to physicians. The x-axis values range between 0 and 20 in order to show clearly over 98% of the observations. In the x-range from 8 to 20, there is no obvious pattern. However, between 0 and 8 we can see more variability, that can be an effect of the skewness of the predicted values. The plot indicates some under-prediction when the predicted number of visits is very small. The results of these preliminary model diagnostics suggest that the use of M-quantile small area estimators is a reasonable choice.

3 M-quantile regression

In this Section we present an extension of linear M-quantile regression to count data following Chambers et al. (2012b) and Chambers et al. (2012a). We start by providing a fairly detailed presentation of M-quantile regression for continuous outcomes before focusing on the case of count outcomes. In this Section we drop subscript d for ease of notation.

3.1 M-quantile regression for a continuous response

The classic regression model summarises the behaviour of the mean of a random variable y at each point in a set of covariates x. This provides a rather incomplete picture, in much the same way as the mean gives an incomplete picture of a distribution. Quantile regression summarises the behaviour of different parts (e.g.

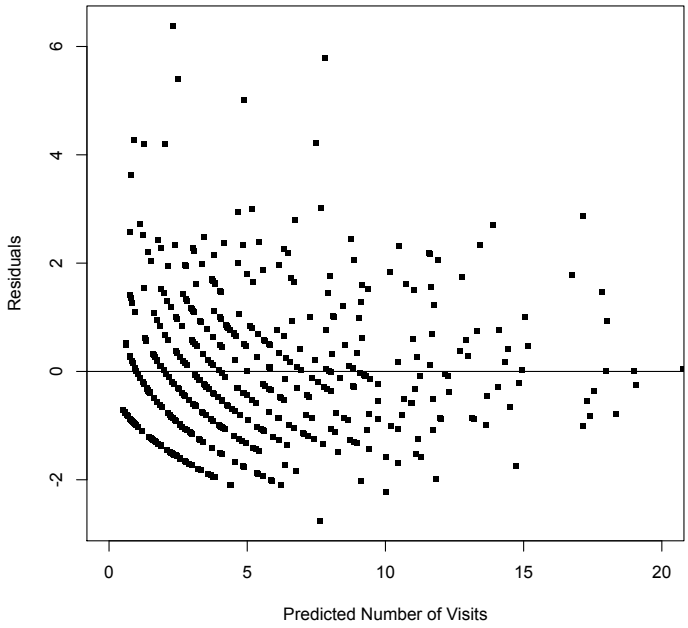

Figure 3: Model fit diagnostics for a Poisson-Normal mixed model fit to the data: raw residuals vs predicted values.

quantiles) of the conditional distribution of y at each point in the set of the x's. In the linear case, quantile regression leads to a family of hyper-planes indexed by a real number $q \in (0, 1)$. For a given value of q, the corresponding model shows how the q-th quantile of the conditional distribution of y varies with x. For example, if $q = 0.5$ the quantile regression hyperplane shows how the median of the conditional distribution changes with x. Similarly, for $q = 0.1$ the quantile regression hyperplane separates the lower 10% of the conditional distribution from the remaining 90%.

Suppose $(\mathbf{x}_j^T, y_j),\ j = 1, \ldots, n$ denotes the values observed for a random sample consisting of n independent observations from a population, where $\mathbf{x}_j^T$ are row p-vectors of a known design matrix $\mathbf{X}$ and y_j is a scalar response variable corresponding to a realisation of a continuous random variable with unknown continuous cumulative distribution function F. A linear regression model for the q-th conditional quantile of y_j given $\mathbf{x}_j$ is

$$Q_y(q|\mathbf{x}_j) = {\mathbf{x}_j}^T \boldsymbol{\beta}_q. \tag{4}$$

An estimate of the q-th regression parameter $\boldsymbol{\beta}_q$ is obtained by minimizing

$$\sum_{j=1}^{n} |y_j - \mathbf{x}_j^T \boldsymbol{\beta}_q| \{(1-q) I(y_j - \mathbf{x}_j^T \boldsymbol{\beta}_q \leq 0) + q I(y_j - \mathbf{x}_j^T \boldsymbol{\beta}_q > 0)\}.$$

Solutions to this problem are usually obtained by linear programming methods (Koenker & D'Orey, 1987) and algorithms for fitting quantile regression are now available in standard statistical software, for example the library `quantreg` in `R` (R Development Core Team, 2010), the command `qreg` in `Stata`, and the procedure `quantreg` in `SAS`.

Quantile regression can be viewed as a generalization of median regression. In the same way, expectile regression (Newey & Powell, 1987) is a 'quantile-like' generalization of mean (i.e. standard) regression. M-quantile regression (Breckling & Chambers, 1988) integrates these concepts within a framework defined by a 'quantile-like' generalization of regression based on influence functions (M-regression). The M-quantile of order q for the conditional density of y given the set of covariates x, $f(y|x)$, is defined as the solution $MQ_y(q|x;\psi)$ of the estimating equation $\int \psi_q\{y - MQ_y(q|x;\psi)\} f(y|x) dy = 0$, where ψ_q denotes an asymmetric influence function, which is the derivative of an asymmetric loss function ρ_q. A linear M-quantile regression model y_j given $\mathbf{x}_j$ is one where we assume that

$$MQ_y(q|\mathbf{x}_j;\psi) = {\mathbf{x}_j}^T \boldsymbol{\beta}_q. \tag{5}$$

That is, we allow a different set of p regression parameters for each value of $q \in$

$(0, 1)$. Estimates of $\boldsymbol{\beta}_q$ are obtained by minimizing

$$\sum_{j=1}^{n} \rho_q(y_j - \mathbf{x}_j{}^T \boldsymbol{\beta}_q). \tag{6}$$

Different regression models can be defined as special cases of (6). In particular, by varying the specifications of the asymmetric loss function ρ_q we obtain the expectile, M-quantile and quantile regression models as special cases. When ρ_q is the squared loss function we obtain the linear expectile regression model if $q \neq 0.5$ (Newey & Powell, 1987) and the standard linear regression model if $q = 0.5$. When ρ_q is the loss function described by (Koenker & Bassett, 1978) we obtain the linear quantile regression.

Setting the first derivative of (6) equal to zero leads to the following estimating equations

$$\sum_{j=1}^{n} \psi_q(r_{jq})\mathbf{x}_j = 0, \tag{7}$$

where $r_{jq} = y_j - \mathbf{x}_j^T \boldsymbol{\beta}_q$, $\psi_q(r_{jq}) = 2\psi(s^{-1}r_{jq})\{qI(r_{jq} > 0) + (1-q)I(r_{jq} \leq 0)\}$ and $s > 0$ is a suitable estimate of scale. For example, in the case of robust regression, $s = \text{median}|r_{jq}|/0.6745$, and we use the Huber Proposal 2 influence function, $\psi(u) = uI(-c \leq u \leq c) + c \cdot \text{sgn}(u)I(|u| > c)$. Provided that the tuning constant c is strictly greater than zero, estimates of $\boldsymbol{\beta}_q$ are obtained using iterative weighted least squares (IWLS).

3.2 M-quantile regression for count data: A Quasi-likelihood approach

The use of M-quantile regression with discrete outcomes is challenging as in this case there is no agreed definition of an M-quantile regression function (Chambers et al., 2012b,a). A popular approach for modelling the mean of a discrete outcome as a function of predictors is via the use of GLMs by assuming that the response variable follows a distribution that is a member of the exponential family of distributions using an appropriate link function. For count data an appropriate distribution is the Poisson and the link function is the logarithm.

In the same way that we impose in the linear specification (4) the continuous case, we impose an appropriate continuous (in q) specification on $MQ_y(q|\mathbf{X}; \psi)$ for count data (Chambers et al., 2012b,a). The most obvious specification for count data is the log-linear specification. That is, we replace (5) by

$$MQ_y(q|\mathbf{x}_j; \psi) = t_j \exp(\mathbf{x}_j \boldsymbol{\beta}_q), \tag{8}$$

where t_j is an offset term. For estimating $\boldsymbol{\beta}_q$, following Chambers et al. (2012b,a), we consider extensions of the robust version of the estimating equations for GLMs by Cantoni & Ronchetti (2001) to the M-quantile case. In particular, Cantoni & Ronchetti (2001) propose a robust version of the estimating equations for GLMs and consider two popular GLMs namely, the binomial and the Poisson models. Estimating equations are defined by

$$\Psi(\boldsymbol{\beta}) := n^{-1}\sum_{j=1}^{n}\Big\{\psi(r_j)w(\mathbf{x}_j)\frac{1}{\sigma(\mu_j)}\mu_j' - a(\boldsymbol{\beta})\Big\} = \mathbf{0}, \tag{9}$$

where $r_j = \sigma(\mu_j)^{-1}(y_j - \mu_j)$ are Pearson residuals, $E[Y_j] = \mu_j$, μ_i' is its derivative with respect to $\boldsymbol{\beta}$, Var$[Y_j] = \sigma^2(\mu_j)$, and $a(\boldsymbol{\beta}) = n^{-1}\sum_{j=1}^{n} E[\psi(r_j)]w(\mathbf{x}_j)\mu_j'/\sigma(\mu_j)$ ensures the Fisher consistency of the estimator. The bounded ψ function is introduced to control deviation in y-space, whereas weights $w(\cdot)$ are used to downweight the leverage points. When $w(\mathbf{x}_j) = 1,\ j = 1,\ldots,n$ Cantoni & Ronchetti (2001) call the estimator the Huber quasi-likelihood estimator. Notice that when ψ is the identity function we obtain the classic quasi-likelihood estimator for GLMs.

For M-quantile regression the estimating equations (9) can be re-written as

$$\Psi(\boldsymbol{\beta}_q) := \frac{1}{n}\sum_{j=1}^{n}\Big\{\psi_q(r_{jq})w(\mathbf{x}_j)\frac{1}{\sigma(MQ_y(q|\mathbf{x}_j;\psi))}MQ_y'(q|\mathbf{x}_j;\psi) - a(\boldsymbol{\beta}_q)\Big\} = \mathbf{0}, \tag{10}$$

where $r_{jq} = \sigma(MQ_y(q|\mathbf{x}_j;\psi))^{-1}(y_j - MQ_y(q|\mathbf{x}_j;\psi))$, $\sigma(MQ_y(q|\mathbf{x}_j;\psi)) = = MQ_y(q|\mathbf{x}_j;\psi)^{1/2}$, $MQ_y'(q|\mathbf{x}_j;\psi) = MQ_y(q|\mathbf{x}_j;\psi)\mathbf{x}_j$ and $a(\boldsymbol{\beta}_q)$ is a correction term to obtain unbiased estimators, which is defined following the arguments in Cantoni & Ronchetti (2001),

$$a(\boldsymbol{\beta}_q) = n^{-1}\sum_{j=1}^{n} 2w_q(r_{jq})w(\mathbf{x}_j)\Big\{cP(Y_j \geqslant i_2+1) - cP(Y_j \leqslant i_1) +$$

$$\frac{MQ_y(q|\mathbf{x}_j;\psi))}{\sigma(MQ_y(q|\mathbf{x}_j;\psi))}[P(Y_j = i_1) - P(Y_j = i_2)]\Big\}MQ_y(q|\mathbf{x}_j;\psi)^{1/2}\mathbf{x}_j,$$

with

- $i_1 = \lfloor MQ_y(q|\mathbf{x}_j;\psi) - c\sigma(MQ_y(q|\mathbf{x}_j;\psi))\rfloor$,
- $i_2 = \lfloor MQ_y(q|\mathbf{x}_j;\psi) + c\sigma(MQ_y(q|\mathbf{x}_j;\psi))\rfloor$ and
- $w_q(r_{jq}) = [qI(r_{jq} > 0) + (1-q)I(r_{jq} \leqslant 0)]$.

When $w(\mathbf{x}_j) = 1, j = 1,\ldots,n$ a Huber quasi-likelihood estimator is obtained. An alternative simple choice for $w(\mathbf{x}_j)$ suggested by robust estimation in linear models is $w(\mathbf{x}_j) = \sqrt{1-h_j}$ where h_j is the jth diagonal element of the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

The solution to the estimating equations (10) can be obtained numerically by a Fisher scoring procedure.

Note that (9) can be obtained as special case of (10) for specific choices of q. In particular, when $q = 0.5$ we obtain (9). Moreover, linear M-quantile regression is a special case of (10) if the the linear link function $MQ_y(q|\mathbf{x}_j;\psi) = \mathbf{x}_j^T\boldsymbol{\beta}_q$ is used and c tends to infinity. `R` routines for fitting M-quantile regression for count data are available from the authors.

3.3 Alternative estimation approaches

Quantile regression for count data from a Bayesian perspective has been recently considered by Lee & Neocleous (2010) whereas from a frequentist perspective by Machado & Santos Silva (2005). In both papers, the authors point out that the problem with estimating conditional quantiles of counts is caused by the combination of a non differentiable sample objective function with a discrete dependent variable. To overcome this problem, Machado & Santos Silva (2005) use a specific form of jittering for creating artificial smoothness in the outcome. In particular, smoothness is achieved by adding to the count outcome noise generated from a Uniform$(0, 1)$. The quantiles of the resulting continuous outcome are then directly modelled because of the one-to-one relationship between the conditional quantiles of the count outcome and those of the artificially generated continuous outcome. An alternative approach to modeling the conditional distribution of a count outcome given the covariates was proposed by Efron (1992) who proposed using asymmetric maximum likelihood estimation. As Machado & Santos Silva (2005) point out, asymmetric maximum likelihood estimation can be seen as the result of smoothing the objective function used to define the quantile regression estimator. Efron's approach results in estimates of conditional location for count data that is similar to conditional expectiles proposed by Newey & Powell (1987).

The approach we propose in this paper for estimating M-quantile regression also uses an objective function that has a degree of smoothness. In particular, the smoothness can be increased by setting the tuning constant in the Huber influence function equal to a large value in which case estimates of the model parameters from our approach should be close to those obtained by Efron's asymmetric maximum likelihood estimation. Indeed, comparing estimates of the model parameters from Efron's (1992) method (using the `vgam` function with family equal to `amlpoisson` in `R`) with our method, when setting Huber's tuning constant equal to a large value, confirm this assumption. Nevertheless, more needs to be done for comparing the different estimation approaches especially when these are used for prediction purposes as is the case with small area estimation.

4 Estimation of small area counts by M-quantile regression models

4.1 Point estimation

Linear mixed effects models and GLMMs include random area effects to account for between-area variation. Estimation of the model parameters is then implemented by means of parametric assumptions such as that random effects are normally distributed. Efficient prediction of random effects is crucial due to their central role in small area estimation. Although for linear models closed form solutions exist, for GLMs this is not the case. For GLMMs and from a frequentist perspective predicted random effects are obtained by using approximations to the likelihood, for example via first or second order penalised quasi-likelihood, or numerical methods such as Gaussian quadrature. Hence, for GLMMs outlier robust prediction of random effects becomes more challenging and the use of semi-parametric methods may offer a simpler solution to outlier robust estimation.

A key concept in the application of M-quantile methods to data with group structure is the identification of a unique 'M-quantile coefficient' associated with each observed datum. These coefficients are then averaged, in some suitable way, over observations making up the group to define a group level M-quantile coefficient, which can be used to characterise the distribution of $y|x$ within the group in very much the same way as a random group effect. In the continuous y case, the M-quantile coefficient for observation j is simply defined as the unique solution q_j to the equation $y_j = \widehat{MQ}_y(q_j|\mathbf{x}_j;\psi)$. However, for count data the equation $y_j = \widehat{MQ}_y(q_j|\mathbf{x}_j;\psi)$ does not have a solution when $y_j = 0$. To overcome this problem we use the definition by Chambers et al. (2012a):

$$\widehat{MQ}_y(q_j|\mathbf{x}_j;\psi) = \begin{cases} k(\mathbf{x}_j) & y_j = 0 \\ y_j & y_j = 1, 2, \ldots \end{cases}$$

A possibility is $k(\mathbf{x}_j) = \widehat{MQ}_y(q_{\min}|\mathbf{x}_j;\psi)$ where $q_{\min}$ denotes the smallest q-value in the grid of q-values used to determine the q_j values of the observed units. However, this implies that $q_j = q_{\min}$ whenever $y_j = 0$, irrespective of the value of $\mathbf{x}_j$, which does not appear to be appropriate. One way to tackle this is by following the same line of argument that Chambers et al. (2012b) used in motivating the definition of q_j for the Bernoulli case. This implies that an observation with value $y_1 = 0$ corresponds to a smaller q-value than another with value $y_2 = 0$ when $\widehat{MQ}_{y_1}(0.5|\mathbf{x}_1;\psi) > \widehat{MQ}_{y_2}(0.5|\mathbf{x}_2;\psi)$. A way to define this is by setting $k(\mathbf{x}_j) = \min\{1-\epsilon, [\widehat{MQ}_y(0.5|\mathbf{x}_j;\psi)]^{-1}\}$, where $\epsilon > 0$ is a small positive constant.

Then the M-quantile coefficient for unit j is q_j, where

$$\widehat{MQ}_y(q_j|\mathbf{x}_j;\psi) = \begin{cases} \min\left\{1-\epsilon, \dfrac{1}{\exp(\mathbf{x}_j^T\hat{\boldsymbol{\beta}}_{0.5})}\right\} & y_j = 0 \\ y_j & y_j = 1, 2, \ldots \end{cases} \tag{11}$$

For a detailed discussion see Chambers et al. (2012a,b).

Provided there are sample observations in area d, an area d specific M-quantile coefficient, $\hat{\theta}_d$ can be defined as the average value of the sample M-quantile coefficients in area d, otherwise we set $\hat{\theta}_d = 0.5$. Following Chambers & Tzavidis (2006), the M-quantile predictor of the average count $\bar{y}_d$ in small area d is then

$$\hat{\bar{y}}_d^{MQ} = N_d^{-1}\Big\{\sum_{j\in s_d} y_{dj} + \sum_{j\in r_d} \widehat{MQ}_y(\hat{\theta}_d|\mathbf{x}_{dj};\psi)\Big\}, \tag{12}$$

where $\widehat{MQ}_y(\hat{\theta}_d|\mathbf{x}_{dj};\psi) = \exp\{\mathbf{x}_{dj}^T\hat{\boldsymbol{\beta}}_{\hat{\theta}_d}\}$.

4.2 Mean squared error estimation

The Mean Squared Error of the predictor $\hat{\bar{y}}_d^{MQ}$ is defined as

$$MSE(\hat{\bar{y}}_d^{MQ}) = E[(\hat{\bar{y}}_d^{MQ} - \bar{y}_d)^2]. \tag{13}$$

Following Chambers et al. (2012b) we propose a nonparametric bootstrap-based estimator of the MSE of the $\hat{\bar{y}}_d^{MQ}$ by constructing an artificial finite population that resembles the real population. To develop the bootstrap procedure we write the linear predictor of the Poisson M-quantile regression model in a form that mimics the mixed effects model form,

$$y_{dj} = \mathbf{x}_{dj}^T\boldsymbol{\beta}_{0.5} + \mathbf{x}_{dj}^T(\boldsymbol{\beta}_{\theta_d} - \boldsymbol{\beta}_{0.5}). \tag{14}$$

Averaging the last term on the right-hand side of (14) for each small area results in a term u_d^{MQ} which can be interpreted as a pseudo-random effect for area d in that it quantifies an average difference of the area-specific M-quantile fit from the median fit.

The bootstrap we propose is nonparametric in nature in the sense that the area effects are generated by using an empirical rather than a parametric distribution. The steps of the bootstrap procedure are summarized below:

- (Step 1) Using sample s, fit model (8) and obtain predictors $\hat{\bar{y}}_d^{MQ}$. For each small area compute the pseudo-random effect $\hat{u}_d^{MQ}$ by computing the $E(\mathbf{x}_{dj}^T(\boldsymbol{\beta}_{\theta_d} - \boldsymbol{\beta}_{0.5}))$ for each area. It is convenient to re-scale the elements $\hat{\mathbf{u}}^{MQ}$ so that they have sample mean exactly equal to zero.

- (Step 2) Construct the vector $\hat{\mathbf{u}}^{MQ*} = \{\hat{u}_1^{MQ*}, \ldots, \hat{u}_D^{MQ*}\}^T$, whose elements are obtained by extracting a simple random sample with replacement of size D from the set $\{\hat{u}_1^{MQ}, \ldots, \hat{u}_D^{MQ}\}^T$.
- (Step 3) Generate a bootstrap population U^* of size $N = \sum_{d=1}^{D} N_d$, by generating values from a poisson distribution with

$$\mu_{dj}^* = \exp\{\mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}}_{0.5} + \hat{u}_d^{MQ*}\},\ j = 1, \ldots, N_d$$

  and calculate the bootstrap population parameters $\bar{y}_d^*$, $d = 1, \ldots, D$. The choice of a Poisson distribution may appear contradictory to the non-parametric nature of the proposed bootstrap. However, the Poisson assumption is implicit in our M-quantile Poisson model due to the mean-variance relationship in the estimating equations. The non-parametric aspect of the proposed bootstrap is only related to the way the pseudo-random effects are generated.
- (Step 4) Extract a sample s^* of size n from the bootstrap population U^* using the assumed sampling design and compute small area estimates with the bootstrap sample $\hat{\bar{y}}_d^{MQ*},\ d = 1, \ldots, D$.
- (Step 5) Repeat steps 2-4 B times.
- (Step 6) Denoting by $\hat{\bar{y}}_d^{MQ*(b)}$ the M-quantile predictor in the b-th bootstrap replication and by $\bar{y}_d^{*(b)}$ the corresponding population value in the b-th bootstrap population, a bootstrap estimator of MSE is

$$\text{MSE}(\hat{\bar{y}}_d^{MQ}) = B^{-1} \sum_{b=1}^{B} \left(\hat{\bar{y}}_d^{MQ*(b)} - \bar{y}_d^{*(b)} \right)^2. \tag{15}$$

The proposed bootstrap is not the only approach to MSE estimation. An alternative approach would have been to use the random effects block bootstrap (Chambers & Chandra, 2012), which is free both of the distribution and the dependence assumptions of the usual parametric bootstrap. Chambers et al. (2012b) adapted the block bootstrap for estimating the MSE of the M-quantile small area predictor in the case of a Bernoulli outcome. A similar approach can be used in the case of a count outcome. A comparison between the alternative approaches to MSE estimation will be discussed in future work.

5 Results and discussion

In this section we present the results of the application of the small area estimation procedure introduced in the previous section to data from the HCAMS. In particular, we compare small area estimates of the average number of visits to physicians

Table 2: Analysis of quasi-deviance table for the Poisson M-quantile model at $q = 0.50$.

Covariates	Resid. df	df	Deviance	p-value
Null	506			
age	502	4	52.473	1.0e-10
age, gender	501	1	10.375	0.0013
age, gender, region	499	2	10.991	0.0041

among elderly (aged 65 or more) and their corresponding mean squared error estimates based on the following estimators: (i) the direct estimator computed as a ratio estimator using calibration weights provided with the micro-data; (ii) the EBPP in equation (3) based on a Poisson-Normal model with random area intercepts using age, gender and region as covariates according the results of Table 1; (iii) the M-quantile predictor in (12) based on the Poisson M-quantile model introduced in Section 3.2.

In the Poisson M-quantile model the ψ function is set to be the Huber Proposal 2 with the tuning constant $c = 1.6$ (Cantoni & Ronchetti, 2001). In addition, model selection is carried out via a robust stepwise procedure based on the Huber quasi-deviance at $q = 0.5$ (Cantoni & Ronchetti, 2001). The analysis of deviance reported in Table 2 shows that the auxiliary variables age, gender and region, added sequentially, are highly significant on the basis of their deviance value. Interactions between pairs of these variables are again nonsignificant. Table 3 reports the estimated $\boldsymbol{\beta}_q$ coefficients at $q = 0.50$, together with standard errors estimated using the proposal in Cantoni & Ronchetti (2001) and p-values. Estimates confirm what expected: the number of visits increases as the people grow old and women go to the physicians more often than men. This latter figure can be explained by the greater longevity of women that is reflected in a larger number of years in conditions of disability or in presence of chronicities.

Efficient estimates of area effects are necessary for small area estimation via GLMMs. Similarly, estimation of M-quantile coefficients is necessary for small area estimation using the Poisson M-quantile model proposed in this paper. Figure 4 shows how the standardized M-quantile coefficients estimated via expression (11) are related to the standardized area effects estimated using the `glmer` function in `R`. Figure 4 shows that the relationship between the estimated area effects and the estimated M-quantile coefficients is strong. The correlation between the estimated area effects and the estimated M-quantile coefficients is 0.91. This result suggests that M-quantile coefficients are comparable to estimated area effects obtained using standard GLMM fitting procedures as far as capturing intra-area (domain) variability is concerned.

Table 3: Estimated Poisson M-quantile $\boldsymbol{\beta}_q$ coefficients and their standard errors at $q = 0.50$. The baseline for variable age is 65-69, for variable gender is female and for variable region is Liguria.

Covariates	Estimate	Std. Error	p-value
Intercept	-0.411	0.047	1.2e-09
age 70-74	0.029	0.052	0.2854
age 75-79	0.245	0.051	1.1e-06
age 80-84	0.211	0.069	0.0011
age >84	0.256	0.064	3.1e-05
gender	-0.125	0.038	0.0005
region Toscana	0.110	0.044	0.0067
region Umbria	0.149	0.046	0.0007

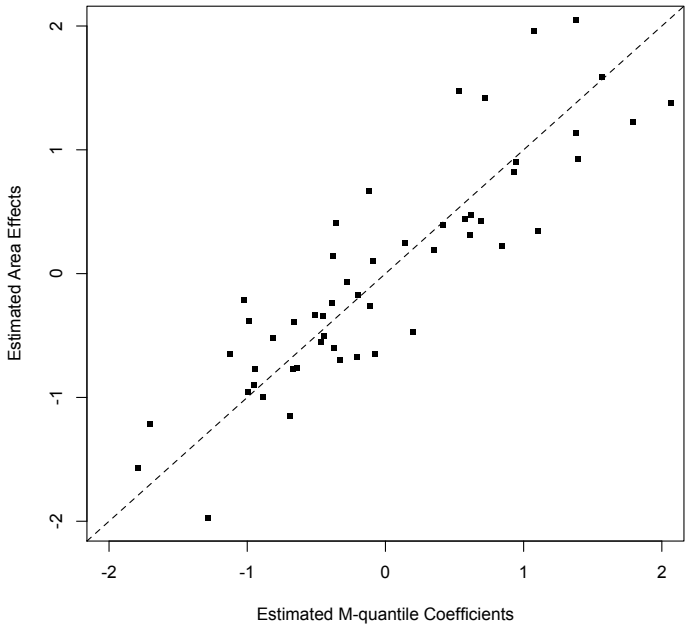

Figure 4: Estimated M-quantile coefficients vs. estimated area effects (standardized values).

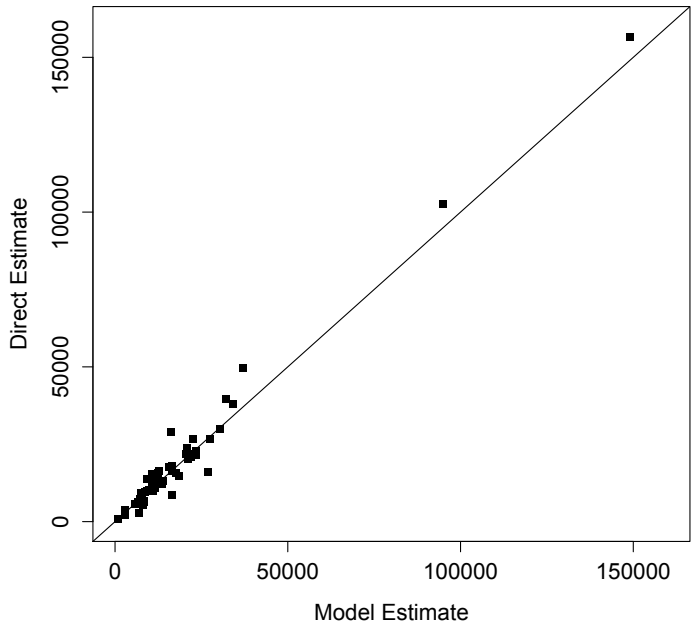

Figure 5: Total number of visits to physicians in Health Districts of Liguria, Toscana and Umbria in 2000: Model-based M-quantile estimates versus corresponding direct estimates.

For comparing the performance of the different estimators we must use a set of diagnostics. Such diagnostics are suggested in Brown et al. (2001). Model-based estimates should be (i) coherent with unbiased direct estimates and (ii) more precise than direct estimates. To validate the reliability of the model-based small-area estimates, we use the goodness of fit (GoF) diagnostic and the values of the coefficient of variation (CV). Overall, the correlation between the model-based estimates and the direct estimates are positive and high, which indicates that the model-based estimates are coherent with the direct estimates (Direct/M-quantile is 0.87 and Direct/EBPP is 0.95). This result for M-quantile estimates is confirmed by Figure 5 where the direct estimates are plotted vs M-quantile estimates: we note that M-quantile estimates appear to be consistent with direct estimates of the total number of visits to physicians.

The GoF diagnostic is based on the null hypothesis that the direct and model-based estimates are statistically equivalent. The alternative is that the direct and model-based estimates are statistically different. The GoF diagnostic is computed

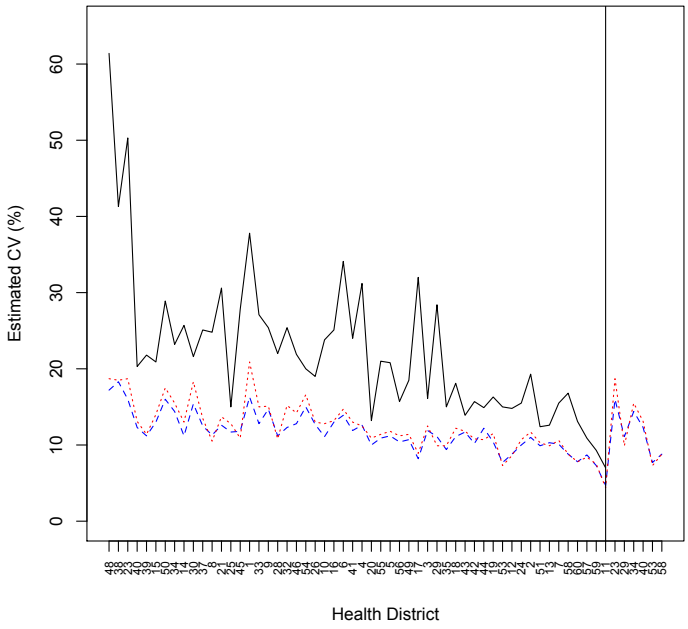

Figure 6: The plot shows distribution of Health Districts values of estimated CV for direct (solid line) and model-based estimates, with estimated CVs for the M-quantile predictor shown as a dashed blue line and estimated CVs for the EBPP shown as a dashed red line. HDs are ordered according to increasing sample size. Out of sample areas are the last six areas.

Ratio/n_i	<24	24-100	101-556
Direct	2.36	1.92	1.52
EBPP	1.07	1.06	1.01

Table 4: Mean values across areas of the ratios of the estimated MSE of the direct and EBPP estimators to the estimated MSE of the Poisson M-quantile estimators grouped by area sample sizes.

using the following Wald statistic for every model based estimator

$$W = \sum_d \left\{ \frac{(\hat{\bar{y}}_d^{\text{direct}} - \hat{\bar{y}}_d^{\text{model}})^2}{[\widehat{\text{var}}(\hat{\bar{y}}_d^{\text{direct}}) + \widehat{\text{mse}}(\hat{\bar{y}}_d^{\text{model}})]} \right\}.$$

The value from the test statistic W is compared against the value from a χ^2 distribution with $D = 54$ degrees of freedom. In our case, this value is 72.15 at 5% level of significance. We use the nonparametric bootstrap algorithm for the M-quantile predictor and the bootstrap procedure proposed by González-Manteiga et al. (2007) for EBPP for estimating the MSE for each small area. Variance estimates for the direct estimator have been computed taking into account the complex two stage design employed for HCAMS. In particular, variance estimates for the estimate of the average number of visits to physicians has been computed separately for each of the three regions to provide a first estimate of the design effect (values between 5.6 and 7.1). Then, following Kish (1987, Section 2.6), design effects have been recomputed to account for the fact that the elderly constitute a subdomain that cuts across PSUs (final deff values for each small area between 0.9 and 1.5). The values of the GoF are 27.3 for M-quantile predictor and 15.9 for EBPP. These results indicate that all model-based estimates are not statistically different from the direct estimates.

Figure 6 shows the distribution across HDs of the estimated CVs (expressed in percentage terms) for direct (solid black line) and model-based estimates (blue denotes M-quantile estimates and red denotes EBPP estimates). The estimated gains of the model-based estimates over the direct estimates are large, particularly for HDs with a small number of sampled units. Generally, the M-quantile estimates have a smaller estimated CV than corresponding EBPP estimates.

Moreover, to evaluate the precision of M-quantile predictor, in Table 5 we report the median values across areas of the ratios of the estimated MSE of the EBPP and direct estimators to the estimated MSE of the WMQ estimator for three groups of areas formed according to area-specific sample sizes. The M-quantile estimator is more efficient than the direct and the EBPP when the sample sizes are small. For large sample sizes the improvements in efficiency of M-quantile predictor are smaller; with $n_i > 100$ M-quantile and EBPP predictors seem to be equivalent: the value of the the ratio is 1.01.

In Figure 7 we compare the maps obtained for the average number of visits to physicians in HDs of Liguria, Toscana and Umbria in 2000 as estimated by direct, Poisson M-quantile and EBPP based estimators. We used the same cut-points to depict the three maps. In line with expectations, most HDs have close levels of average number of visits to physicians, but there are areas deviating from the bulk of the distribution in both directions.

As anticipated by the aforementioned correlation coefficients, the estimates are all comparable in magnitude over the three maps. However, maps based on the two model based estimators look very much alike and show a smoother pattern as opposed to that based on design based estimates. In addition, model based maps allow to have estimates also for 'empty' small areas for which the design based estimator cannot be computed. However, when comparing model based maps, we can note that some HDs tend to have a larger estimate in the EBPP based map than that shown in the Poisson M-quantile map. This is due to the fact that those are the small areas with particularly large values of y for which M-quantile models provide more robustness in the final estimates.

6 Simulation study

The purposes of this simulation experiment are: (i) to compare the performance of the M-quantile predictor with that obtained by EBPP and the direct estimator; (ii) to evaluate the performance of the bootstrap mean squared estimator (15) proposed in Subsection 4.2.

The simulated data are generated using the individual $\mathbf{x}_{dj}^T$ values and the estimates of $\hat{\boldsymbol{\beta}} = (-0.44, 0.05, 0.28, 0.27, 0.29, -0.13, 0.16, 0.14)$ and of the standard error $\hat{\varphi} = 0.192$ obtained fitting the GLMM on the real data of Section 5. In each run of the simulation, y values are generated for the groups given by cross-classifying gender by age class for each of $D = 54$ small areas for which we have sample values. In total, we have ten groups for each small area. The value of the y variable for each cell y_{dj} ($d = 1, \ldots, D,\ j = 1, \ldots, 10$) is calculated as Poisson(μ_{dj}) with $\mu_{dj} = N_{dj} \exp\{\eta_{dj}\}$ and $\eta_{dj} = \mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}} + u_d$, where u_d are independently drawn from a normal distribution with mean 0 and standard error $\hat{\varphi}$. Here, $T = 1,000$ populations are generated and the true values of the average number of visits for each of the 54 sampled HDs of Liguria, Toscana and Umbria, $\bar{y}_d = N_d^{-1} \sum_{j=1}^{10} y_{dj}$, $d = 1, \ldots, D$, of the synthetic populations are available.

For each population, sample values y_{dj} for each cell are generated from a Poisson(μ_{dj}^*) with $\mu_{dj}^* = n_{dj} \exp\{\mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}} + u_d\}$, where u_d is the value of random effect drawn previously to create the population, and according to two scenarios:

- (0) - No outliers.

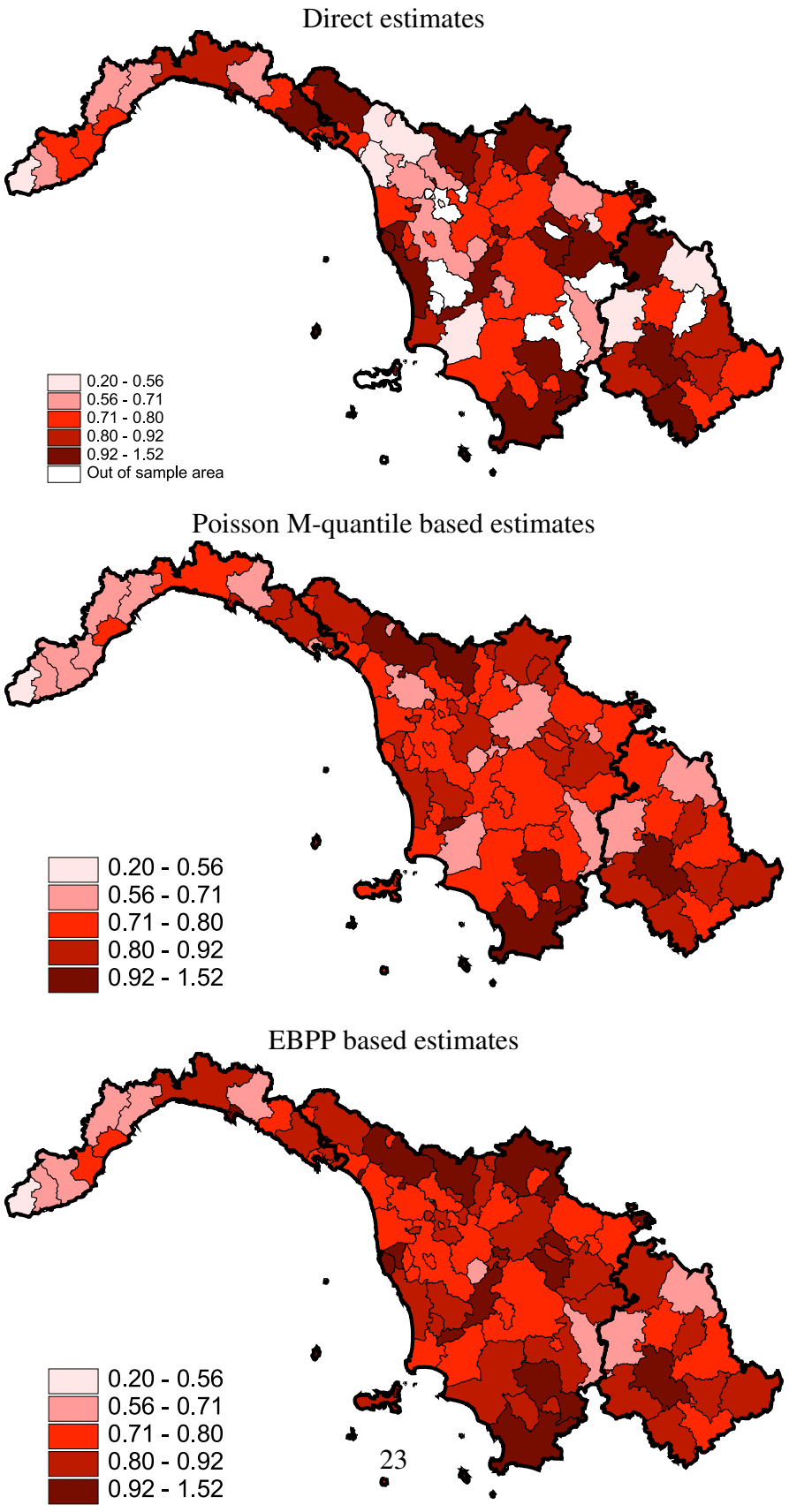

Figure 7: Maps of the direct, Poisson M-quantile and EBPP based estimates of average number of visits to physicians in Health Districts of Liguria, Toscana and Umbria in 2000.

- (M) - Measurement-type error: $2\%, 5\%, 10\%$ of randomly chosen response values has been changed from y_{dj} to $y_{dj} = y_{dj} + 10$.

For each sample the M-quantile, EBPP and the direct estimator are used to estimate the average small area count $\bar{y}_d,\ d = 1, \ldots, D$. The performances of different small area estimators are evaluated with respect to two basic criteria: the bias and the root mean squared error (RMSE). Simulated values of the bias and of the mean squared error for a small area estimator are obtained as $T^{-1}\sum_{t=1}^{T}(\hat{\bar{y}}_{dt} - \bar{y}_{dt})$ and $T^{-1}\sum_{t=1}^{T}(\hat{\bar{y}}_{dt} - \bar{y}_{dt})^2$, respectively. Here $\bar{y}_{dt}$ denotes the actual area d value at simulation t, with predicted value $\hat{\bar{y}}_{dt}$. The median and maximum value of the absolute Bias and the median value of RMSE over small areas are set out in Table 5. The results confirm our expectations regarding the behaviour of the estimators: under the (0) scenario the EBPP performs slightly better than the M-quantile in terms of RMSE, whereas there is no noticeable difference among the three estimators in terms of bias. The M-quantile predictor is the best in terms of bias and RMSE under the (M) scenarios and it is clearly superior respect to the other estimators as contamination increases.

Table 5: Model-based simulation results: performances of predictors of small area count. Scenarios (0) and (M), Contamination: $2\%, 5\%, 10\%$, $D = 54$.

Predictor/Scenario	(0)	(M) 2%	(M) 5%	(M) 10%
	Median values of Absolute Bias			
EBPP	0.0024	0.0360	0.0863	0.1814
M-quantile	0.0085	0.0095	0.0299	0.0699
Direct	0.0147	0.0318	0.0828	0.1785
	Maximum value of Absolute Bias			
EBPP	0.0809	0.1050	0.1356	0.1974
M-quantile	0.0112	0.0775	0.1826	0.3837
Direct	0.0970	0.1431	0.2891	0.5553
	Median values of RMSE			
EBPP	0.0973	0.1217	0.1624	0.2548
M-quantile	0.1072	0.1088	0.1163	0.1410
Direct	0.1272	0.1562	0.1966	0.2790

Regarding the second purpose of the simulation study, i.e. the evaluation of the performance of the bootstrap MSE estimator (15) proposed in Section 4.2, we use the data generated for scenarios (0) and (M)-10% and a subset of small areas: $D = 14$, the HDs of the Region Liguria. The results of the MSE estimator, based on 500 bootstrap iterations, for each scenario are shown in Table 6 where we report the median values of their area specific biases and their root mean squared errors,

expressed in relative terms (%). The MSE estimator shows small bias and a good stability under both scenarios. In particular, under scenario (M)-10% tend to be biased up and its Relative RMSE increases not much respect the no-contaminated scenario. Examination of Table 6 shows that MSE estimation method generate nominal 95 per cent confidence intervals with a small under coverage. A bootstrap bias correction term could be included in expression (15). This term will be a part of a future research.

Table 6: Model-based simulation results: performance of bootstrap MSE estimator (15). Scenarios (0) and (M)-10%, $D = 14$, Median values.

Indicator/Scenario	(0)	(M) 10%
Relative bias	-3.05	1.61
Relative RMSE	27.44	31.09
Coverage rate (95% nominal)	90%	86%

7 Final remarks

To carry on SAE for count data, Poisson M-quantile models are introduced and investigated. Such models offer a natural way of modeling between-area variability in the data, without imposing prior assumptions on the source of this variability or a pre-specified hierarchical structure. 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 is straightforward under these models.

The proposed approach looks suitable for estimating a wide range of parameters and the model-based simulation suggests that it is a reasonable alternative to mixed effects models. In fact, M-quantile regression-based methods outperform GLMM-based methods when data include outliers. And, in case there are no outliers, there is no noticeable loss in efficiency.

The occurrence of this type of data in real situations, as well as the usefulness of the suggested approach, have been described by a real application. The model allows us to estimate the average number of recourse to physicians for each HDs, stratified for age and gender, for three Italian regions with older population.

Even though the suggested approach provides encouraging results, further investigation is still necessary. A first topic of future research is the choice of the

area quantile coefficients, trying alternative estimation of q_{dj}. Second, an analytical estimator of MSE (instead of a nonparametric bootstrap estimator) could be attempted. Third, because of additivity of the Poisson distribution, on a small area estimation perspective, benchmarking at different size level could be achieved.

Another important issue is how to take overdispersion on data into account. One possibility could be using Quasi-Poisson M-quantile models. Another option is Negative Binomial M-quantile models (see Chambers et al., 2012a). In the real application considered in this paper, Poisson M-quantile models seem to be able to account for overdispersion. In fact, Poisson and Quasi-Poisson M-quantile models provide comparable results. So, it could be interesting to evaluate when Quasi-Poisson models are really necessary in real applications.

Finally, other newsworthy topics for future research are: (i) to develop zero-inflated M-quantile Poisson regression in order to take the excess of zero on the data into account; (ii) to adapt the suggested model to complex sampling designs.

References

BATTESE, G., HARTER, R. & FULLER, W. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* **83**, 28–36.

BRECKLING, J. & CHAMBERS, R. (1988). M-quantiles. *Biometrika* **75**, 761–771.

BROWN, G., CHAMBERS, R., HEADY, P. & HEASMAN, D. (2001). *Evaluation of small area estimation methods - An application to unemployment estimates from the UK LFS*. Proceedings of Statistics Canada Symposium 2001. Achieving Data Quality in a Statistical Agency: A Methodological Perspective.

CANTONI, E. & RONCHETTI, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association* **96**, 1022–1030.

CHAMBERS, R. & CHANDRA, H. (2012). A random effect block bootstrap for clustered data. *Journal of Computational and Graphical Statistics* , To Appear.

CHAMBERS, R., CHANDRA, H., SALVATI, N. & TZAVIDIS, N. (2013). Outlier robust small area estimation. *To appear in the Journal of the Royal Statistical Society Series B* **75**.

CHAMBERS, R., DREASSI, E. & SALVATI, N. (2012a). Disease mapping via negative binomial m-quantile regression. *Submitted: available from the authors upon request* .

CHAMBERS, R., SALVATI, N. & TZAVIDIS, N. (2012b). M-quantile regression models for binary data in small area estimation. *Submitted: available from the authors upon request* .

CHAMBERS, R. & TZAVIDIS, N. (2006). M-quantile models for small area estimation. *Biometrika* **93**, 255–268.

EFRON, B. (1992). Poisson overdispersion estimates based on the method of asymmetric maximum likelihood. *Journal of the American Statistical Association* **87**, 98–107.

GONZÁLEZ-MANTEIGA, W., LOMBARDÍA, M., MOLINA, I., MORALES, D. & SANTAMARÍA, L. (2007). Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Computational Statistics and Data Analysis* **51**, 2720–2733.

JIANG, J. (2003). Empirical best prediction for small-area inference based on generalized linear mixed models. *Journal of Statistical Planning and Inference* **111**, 117–127.

JIANG, J. & LAHIRI, P. (2001). Empirical best prediction for small area inference with binary data. *Ann. Inst. Statist. Math.* **53**, 217–243.

JIANG, J. & LAHIRI, P. (2006). Mixed model prediction and small area estimation. *TEST* **15**, 1–96.

KISH, L. (1987). *Statistical design for research*. Wiley Series in Probability and Mathematical Statistics.

KOENKER, R. & BASSETT, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

KOENKER, R. & D'OREY, V. (1987). Computing regression quantiles. *Biometrika* **93**, 255–268.

LEE, D. & NEOCLEOUS, T. (2010). Bayesian quantile regression for count data with application to environmental epidemiology. *Journal of the Royal Statistical Society Series C* **59**, 905–920.

MACHADO, J. & SANTOS SILVA, J. M. C. (2005). Quantiles for counts. *Journal of the American Statistical Association* **100**, 1226–1237.

MAITI, T. (2001). Robust generalized linear mixed models for small area estimation. *Journal of Statistical Planning and Inference* **98**, 225–238.

MALEC, D., SEDRANSK, J., MORIARITY, C. & LECLERE, F. (1997). Small area inference for binary variables in the national health interview survey. *Journal of the American Statistical Association* **92**, 815–826.

MC CULLOCH, C. E. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association* **89**, 330–335.

MC CULLOCH, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170.

MOLINA, I. & RAO, J. N. K. (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics* **38**, 369–385.

NEWEY, W. & POWELL, J. (1987). Asymmetric least squares estimation and

testing. *Econometrica* **55**, 819–847.

R DEVELOPMENT CORE TEAM (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

RAO, J. N. K. (2003). *Small Area Estimation*. John Wiley & Sons.

SAEI, A. & CHAMBERS, R. (2003). Small area estimation under linear and generalized linear mixed models with time and area effects. In *S3RI Methodology Working Papers*. Southampton: ed. Southampton Statistical Sciences Research Institute, pp. 1–35.

SINHA, S. (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association* **99**, 451–460.

SINHA, S. & RAO, J. (2009). Robust small area estimation. *Canadian Journal of Statistics* **37**, 381–399.

SONG, P. X., FAN, Y. & KALBFLEISCH, J. (2005). Maximization by parts in likelihood inference (with discussion). *Journal of the American Statistical Association* **100**, 1145–1158.