Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

12-12

M-Quantile Regression for Binary Data with Application to Small Area Estimation

Ray Chambers, Nicola Salvati and Nikos Tzavidis

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Biometrika (2012), **xx**, x, *pp*. 1–24

Printed in Great Britain

M-Quantile Regression for Binary Data with Application to Small Area Estimation

BY RAY CHAMBERS

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

NICOLA SALVATI

Dipartimento di Statistica e Matematica Applicata all'Economia,
Università di Pisa, Via Ridolfi, 10 - I56124 Pisa, Italy, salvati@ec.unipi.it

NIKOS TZAVIDIS

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

SUMMARY

M-quantile regression models are a robust and flexible alternative to random effects models, particularly in small area estimation. However quantiles, and more generally M-quantiles, are only uniquely defined for continuous variables. In this paper we extend the M-quantile regression approach to binary data, and more generally to count data. This approach is then applied to estimation of a small area proportion, where a popular alternative approach is to use a plug-in version of the Empirical Best (EB) predictor based on a generalised linear mixed model for the underlying binary variable. Results from both model-based and design-based simulations comparing the binary M-quantile and the plug-in EB predictors demonstrate the usefulness of the M-quantile approach in this case. The paper concludes with two illustrative applications. The first addresses estimation of the number of unemployed people aged 16 and above resident in the Unitary Authorities and Local Authority Districts of Great Britain. The second considers estimation of the number of poor households in each of the Local Labour Systems of the Tuscany region of Italy.

Some key words: Influence function; M-estimation; Bootstrap methods; Simulation experiments.

1. INTRODUCTION

The increasing demand for reliable small area statistics has led to the development of a number of efficient model-based small area estimation (SAE) methods (Rao, 2003; Jiang & Lahiri, 2006). 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; Prasad & Rao, 1990). However, using a mixed model to characterise differences between small areas requires strong distributional assumptions, a formal specification of the random part of the model and does not easily allow for outlier robust inference. An alternative approach is to use M-quantile models (Breckling & Chambers, 1988) to characterise these differences (Chambers & Tzavidis, 2006). Unlike traditional mixed models, M-quantile models do not depend on strong distributional assumptions and automatically provide outlier robust inference.

Unfortunately, many survey variables are categorical in nature and are therefore not suited to standard SAE methods based on LMMs. One option in such cases is to adopt a Hierarchical Bayes approach (Malec et al., 1997; Nandram et al., 1999) or to use Empirical Bayes (MacGibbon & Tomberlin, 1989; Farrell et al., 1997). Alternatively, if a frequentist approach is preferred, 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 generalised linear mixed models (GLMMs). However, application of these models in SAE is not straightforward, as computing the maximum likelihood estimates of the fixed and random effects parameters of a GLMM can require evaluation of high dimensional integrals. Furthermore, the EBP is model dependent and 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 generalised linear models (GLMs). This has lead to the development of robust alternative methods for fitting these models (Pregibon, 1982; Preisser & Qaqish, 1999; Cantoni & Ronchetti, 2001). But development of corresponding robust methods for GLMM-based SAE is limited. Maiti (2001) describes a Hierarchical Bayes approach to fitting a GLMM based on an outlier-robust normal mixture prior for the random effects, while 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. Similarly, Noh & Lee (2007) propose methods that allow robust inferences for GLMs, with extension to GLMMs. However, only Maiti (2001) applies these methods to SAE.

To the best of our knowledge, there is no existing theory for robust SAE based on GLMMs in the frequentist framework. In this paper we therefore propose a new approach to SAE for discrete data based on M-quantile modelling. This allows straightforward extension of the existing M-quantile approach for continuous data to the case where the response is binary or, more generally, a count. As with M-quantile modelling 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. Outlier-robust inference is also automatic in case of both misclassification error and measurement error.

In the next Section we summarise estimation of a small area proportion based on a GLMM. In Section 3 we then extend the M-quantile regression approach to binary data and, more generally, to count data. In Section 4 we build on the extension to binary data to define the corresponding M-quantile coefficients of the sample units. As with the continuous case, these coefficients capture small area effects in the data, leading to an M-quantile estimator of a small area proportion. In the same Section, we also propose three estimators of the mean squared error (MSE) of this estimator: an analytical estimator based on first order approximations to the variances of solutions to estimating equations, and two bootstrap methods that extend ideas set out in Tzavidis et al. (2010) and Chambers & Chandra (2012). In Section 5, we present results from model-based and design-based simulation studies aimed at assessing the performance of the different small area predictors considered in this paper. In Section 6 we use the M-quantile approach described in this paper to estimate: (1) the number of unemployed people aged 16 and above resident in each of 406 Unitary Authorities and Local Authority Districts (UALAD's) of the UK; and (2) the number of poor households in each of the 57 Local Labour Systems (LLS's) of the Tuscany region of Italy. Finally, in Section 7 we conclude the paper with a discussion of our results and outstanding research issues.

2. SMALL AREA ESTIMATION BASED ON GENERALISED LINEAR MIXED MODELS

Let U denote a finite population of size N which can be partitioned into D domains or small areas, with U_d denoting small area d. The small area population sizes $N_d; d = 1, ..., D$ are assumed known. Let y_{dj} be the value of the variable of interest (typically 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. The aim is to use the sample values of y_{dj} and the population values of $\mathbf{x}_{dj}$ and $\mathbf{z}_d$ to infer the values $\theta_d; d = 1, ..., D$ of a small area characteristic of interest. To save notation, in what follows we use E_s to denote the expectation conditional on this information. It is well known that the minimum mean squared error predictor of θ_d is then $E_s[\theta_d]$.

In many cases $\theta_d = N_d^{-1} \sum_{j \in U_d} f(y_{dj})$ where f is a known function. The minimum mean squared error predictor of θ_d is then $N_d^{-1}\{\sum_{j \in s_d} f(y_{dj}) + \sum_{j \in r_d} E_s[f(y_{dj})]\}$, where s_d denotes the n_d sampled units in small area d and r_d denotes the $N_d - n_d$ remaining (i.e. non-sampled) units in this area. In general, the conditional expectation $E_s[f(y_{dj})]$ can be difficult to evaluate, and so is replaced by a suitable approximation. One such approximation is $E[f(y_{dj})|\mathbf{u}_d]$ where the $\mathbf{u}_d; d = 1, ..., D$ are q-dimensional independent random effects characterising the between-area differences in the distribution of y_{dj} given $\mathbf{x}_{dj}$ (see Rao, 2003; Jiang & Lahiri, 2006; González-Manteiga et al., 2007). This can be formalised by assuming a generalised linear mixed model (GLMM) for $\mu_{dj} = E[y_{dj}|\mathbf{u}_d]$ of the form

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

where g is a known invertible link function. When y_{dj} is binary-valued a popular choice for g is the logistic link function and the individual y_{dj} values in area d are taken to be independent Bernoulli outcomes with

$$\mu_{dj} = E[y_{dj}|\mathbf{u}_d] = P(y_{dj} = 1|\mathbf{u}_d) = \exp\{\eta_{dj}\}(1 + \exp\{\eta_{dj}\})^{-1} \qquad (2)$$

and $Var[y_{dj}|\mathbf{u}_d] = \mu_{dj}(1 - \mu_{dj})$. The q-dimensional vector $\mathbf{u}_d$ is generally assumed to be independently distributed between areas and to follow a normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_u$. The matrix $\boldsymbol{\Sigma}_u$ is allowed to depend on parameters $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)$, which are then referred to as the variance components of the GLMM, while the vector $\boldsymbol{\beta}$ in (1) is referred to as the fixed effects parameter of this model.

We focus on the situation where the target of inference is the small area d proportion, $\bar{y}_d = N_d^{-1} \sum_{j \in U_d} y_{dj}$ and the Bernoulli-Logistic GLMM (2) is assumed. In this case 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 Empirical Best Predictor (EBP) for the area d proportion $\bar{y}_d$ under (2),

$$\hat{\bar{y}}_d^{EBP} = N_d^{-1} \Big\{ \sum_{j \in s_d} y_{dj} + \sum_{j \in r_d} \hat{\mu}_{dj} \Big\}, \qquad (3)$$

where $\hat{\mu}_{dj} = \exp\{\hat{\eta}_{dj}\}(1 + \exp\{\hat{\eta}_{dj}\})^{-1}$, $\hat{\eta}_{dj} = \mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_d^T \hat{\mathbf{u}}_d$, $\hat{\boldsymbol{\beta}}$ is the vector of the estimated fixed effects and $\hat{\mathbf{u}}_d$ denotes the vector of the estimated area-specific random effects. In the simplest case, $q = 1$ and $\mathbf{z}_d = 1$, in which case the $\mathbf{u}_d$ are scalar small area effects. We refer to (3) in this case as a 'random intercepts' EBP. For more details on this predictor, including estimation of its MSE, see Saei & Chambers (2003), Jiang & Lahiri (2006) and González-Manteiga et al. (2007).

Unfortunately, despite their attractive properties as far as modelling non-normal response variables are concerned, application of GLMMs in small area estimation is not always straightforward. 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 in (2) can be carried out using an iterative procedure that combines Maximum Penalized Quasi-Likelihood (MPQL) estimation of $\boldsymbol{\beta}$ and $\mathbf{u}_d$ with REML estimation of $\boldsymbol{\delta}$. See Saei & Chambers (2003). In the empirical results reported in Section 5, we used `glmer` for parameter estimation.

3. M-QUANTILE REGRESSION MODELS FOR BINARY DATA

In this Section we develop an extension of linear M-quantile regression to binary data. Since M-quantile models do not depend on how areas are specified, we also drop the subscript d in this Section. We start by summarising M-quantile regression for a continuous response.

3·1. *M-quantile regression for a continuous response*

M-quantile regression (Breckling & Chambers, 1988) is a 'quantile-like' generalisation of regression based on influence functions (M-regression). The M-quantile of order q of a continuous random variable Y with distribution function $F(Y)$ is the value Q_q that satisfies

$$\int \psi_q\Big(\frac{Y-Q_q}{\sigma_q}\Big)dF(Y)=0, \tag{4}$$

where $\psi_q(t)=2\psi(t)\{qI(t>0)+(1-q)I(t\leqslant 0)\}$ and ψ is a user-defined influence function. Here σ_q is a suitable measure of the scale of the random variable $Y-Q_q$. Note that when $\psi(t)=t$ we obtain the expectile of order q, which represents a quantile-like generalisation of the mean (Newey & Powell, 1987), and when $\psi(t)=sgn(t)$ we obtain the standard quantile of order q (Koenker & Bassett, 1978).

Breckling & Chambers (1988) define a linear M-quantile regression model as one where the ψ-based M-quantile $Q_q(\mathbf{X};\psi)$ of order q of the conditional distribution of y given the vector of p auxiliary variables $\mathbf{X}$ satisfies

$$Q_q(\mathbf{X};\psi)=\mathbf{X}\boldsymbol{\beta}_q. \tag{5}$$

Let $(y_j,\mathbf{x}_j;j=1,\ldots,n)$ denote the available data. For specified q and continuous ψ, an estimate $\hat{\boldsymbol{\beta}}_q$ of $\boldsymbol{\beta}_q$ is obtained by solving the estimating equation

$$n^{-1}\sum_{i=1}^{n}\psi_q(r_{jq})\mathbf{x}_j=\mathbf{0}, \tag{6}$$

where $r_{jq}=y_j-Q_q(\mathbf{x}_j;\psi)$, $\psi_q(r_{jq})=2\psi(\hat{\sigma}_q^{-1}r_{jq})\{qI(r_{jq}>0)+(1-q)I(r_{jq}\leqslant 0)\}$ and $\hat{\sigma}_q$ is a suitable robust estimator of scale, i.e. $\hat{\sigma}_q=median|r_{jq}|/0.6745$. In this paper we will always use the Huber Proposal 2 influence function $\psi(t)=tI(-c<t<c)+c\cdot sgn(t)I(|t|\geqslant c)$. Provided the tuning constant c is bounded away from zero, we can solve (6) using standard iteratively re-weighted least squares (IRLS).

3·2. *M-quantile regression for a binary response: an estimating equation approach*

There is no obvious definition of a quantile regression function when Y is binary since the order q quantile of a binary variable is not unique. However, provided the underlying influence function ψ is continuous and monotone non-decreasing, the M-quantiles of a binary variable do exist and are unique. This is easily seen by considering the solution to (4) when Y is binary, with $P(Y=1)=p$. In this case (4) becomes

$$pq\psi\left(\frac{1-Q_q}{\sigma_q}\right)=(1-p)\,(1-q)\,\psi\left(\frac{Q_q}{\sigma_q}\right).$$

It is easy to see that when $\psi(t)=t$ and $q=0.5$, the solution to this estimating equation is $Q_{0.5}=p$, as should be the case. Furthermore, when both p and q lie strictly between 0 and 1, the preceding assumptions about ψ ensure that Q_q also lies strictly between 0 and 1 and is monotone non-decreasing in q for fixed p. It is also monotone non-decreasing in p for fixed q under the assumption of a fixed scale parameter. A proof of this is available from the authors on request.

In the same way that we impose a linear specification (5) on $Q_q(\mathbf{X};\psi)$ in the continuous case, we can impose an appropriate continuous (in q) specification on $Q_q(\mathbf{X};\psi)$ in the binary case. In particular, it seems sensible to replace (5) by the linear logistic specification

$$Q_q(\mathbf{x}_j;\psi)=\frac{\exp(\mathbf{x}_j^T\boldsymbol{\beta}_q)}{1+\exp(\mathbf{x}_j^T\boldsymbol{\beta}_q)}. \tag{7}$$

In order to estimate the parameter $\boldsymbol{\beta}_q$ we consider the extension to the M-quantile case of the Cantoni & Ronchetti (2001) approach to robust estimation of the parameters of a GLM. In particular these authors propose a robustified version of the maximum likelihood estimating equations for a GLM of the form:

$$\Psi(\boldsymbol{\beta}):=n^{-1}\sum_{j=1}^{n}\Big\{\psi(r_j)w(\mathbf{x}_j)\frac{1}{\sigma(\mu_j)}\mu_j'-a(\boldsymbol{\beta})\Big\}=\mathbf{0}, \tag{8}$$

where $r_j=\frac{y_j-\mu_j}{\sigma(\mu_j)}$ are the Pearson residuals, $E[Y_j]=\mu_j$, $Var[Y_j]=\sigma^2(\mu_j)$, μ_i' is the derivative of μ_j with respect to $\boldsymbol{\beta}$ and $a(\boldsymbol{\beta})=\frac{1}{n}\sum_{j=1}^{n}E[\psi(r_j)]w(\mathbf{x}_j)\frac{1}{\sigma(\mu_j)}\mu_j'$ ensures the Fisher consistency of the solution to (8). The bounded influence function ψ is used to control outliers in y, whereas the weights w are used to downweight the leverage points. When $w(\mathbf{x}_j)=1\,\forall\,j$ Cantoni & Ronchetti (2001) refer to the solution to (8) as the Huber quasi-likelihood estimator. When ψ is the identity function, (8) reduces to the usual maximum likelihood estimating equations for a GLM.

Under the M-quantile framework the estimating equations (8) can be re-written as

$$\Psi(\boldsymbol{\beta}_q):=n^{-1}\sum_{j=1}^{n}\Big\{\psi_q(r_{jq})w(\mathbf{x}_j)\frac{1}{\sigma(Q_q(\mathbf{x}_j;\psi))}Q_q'(\mathbf{x}_j;\psi)-a(\boldsymbol{\beta}_q)\Big\}=\mathbf{0}, \tag{9}$$

where $r_{jq}=\frac{y_j-Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}$, $\sigma(Q_q(\mathbf{x}_j;\psi))=[Q_q(\mathbf{x}_j;\psi)(1-Q_q(\mathbf{x}_j;\psi))]^{1/2}$, $Q_q'(\mathbf{x}_j;\psi)=\sigma^2(Q_q(\mathbf{x}_j;\psi))\mathbf{x}_j$ and $a(\boldsymbol{\beta}_q)$ is a bias correction term:

$$a(\boldsymbol{\beta}_q)=\Big\{\psi_q\Big(\frac{1-Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}\Big)Q_q(\mathbf{x}_j;\psi)-\psi_q\Big(\frac{Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}\Big)(1-Q_q(\mathbf{x}_j;\psi))\Big\}.$$

Setting $w(\mathbf{x}_j)=1\ \forall\ j$ leads to a Huber quasi-likelihood M-quantile estimator. An alternative choice is $w(\mathbf{x}_j)=\sqrt{1-h_j}$ where h_j is the jth diagonal element of the hat matrix $\mathbf{H}=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. This leads to a Mallows type M-quantile estimator. In either case, the estimating equations (9) can be solved numerically using a Fisher scoring procedure to obtain an estimate $\hat{\boldsymbol{\beta}}_q$ of $\boldsymbol{\beta}_q$.

Note that when $q = 0.5$, (9) reduces to (8). Moreover, (6) is a special case of (9) if the the linear link function $Q_q(\mathbf{x}_j;\psi) = \mathbf{x}_j^T\boldsymbol{\beta}_q$ is used and the tuning constant c in the Huber influence function tends to infinity (i.e. ψ is the identity function).

This estimating equation approach applies quite generally. For example, it can be used when Y is a Poisson random variable. In this case the most appealing specification for the M-quantiles of the conditional distribution of Y is log-linear. That is,

$$Q_q(\mathbf{x}_j;\psi) = k\exp(\mathbf{x}_j\boldsymbol{\beta}_q), \tag{10}$$

where k is an offset term. The parameter $\boldsymbol{\beta}_q$ can then be estimated by solving (9) with $\sigma(Q_q(\mathbf{x}_j;\psi)) = Q_q(\mathbf{x}_j;\psi)^{1/2}$, $Q_q'(\mathbf{x}_j;\psi) = Q_q(\mathbf{x}_j;\psi)\mathbf{x}_j$ and

$$a(\boldsymbol{\beta}_q) = 2w_q(r_{jq})\Big\{cP(Y_j \geqslant i_2+1) - cP(Y_j \leqslant i_1) + \frac{Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}[P(Y_j = i_1) - P(Y_j = i_2)]\Big\},$$

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

Assuming that ψ is a continuous monotone non-decreasing function, we can write down a first order sandwich-type approximation to the variance of (9) of the form

$$Var(\hat{\boldsymbol{\beta}}_q) = n^{-1}\Big\{E\Big[\frac{\partial\Psi(\boldsymbol{\beta}_q)}{\partial\boldsymbol{\beta}_q}\Big]\Big\}^{-1} Var\{\Psi(\boldsymbol{\beta}_q)\}\Big[\Big\{E\Big[\frac{\partial\Psi(\boldsymbol{\beta}_q)}{\partial\boldsymbol{\beta}_q}\Big]\Big\}^{-1}\Big]^T. \tag{11}$$

Here

$$Var\{\Psi(\boldsymbol{\beta}_q)\} = n^{-1}\sum_{j=1}^{n}\Big\{\mathbf{x}_j\sigma^2(Q_q(\mathbf{x}_j;\psi))E\Big[\psi_q^2\Big\{\frac{y_j - Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}\Big\}\Big]\mathbf{x}_j^T\Big\} - \sum_{j=1}^{n} a_j^2(\boldsymbol{\beta}_q),$$

where

$$E\Big[\psi_q^2\Big\{\frac{y_j - Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}\Big\}\Big] = \Big\{\psi_q^2\Big(\frac{1 - Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}\Big)Q_q(\mathbf{x}_j;\psi) + \psi_q^2\Big(\frac{-Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}\Big)(1 - Q_q(\mathbf{x}_j;\psi))\Big\},$$

$a_j^2(\boldsymbol{\beta}_q)$ is the square of the bias correction term for unit j, and the expectation $E\Big[\frac{\partial\Psi(\boldsymbol{\beta}_q)}{\partial\boldsymbol{\beta}_q}\Big]$ is

$$\mathbf{B}(\boldsymbol{\beta}_q) = -n^{-1}\sum_{j=1}^{n}\sigma(Q_q(\mathbf{x}_j;\psi))\Big\{\Big\{\psi_q\Big(\frac{1 - Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}\Big) + \psi_q\Big(\frac{Q_q(\mathbf{x}_j;\psi)}{\sigma(Q_q(\mathbf{x}_j;\psi))}\Big)\Big\}\sigma^2(Q_q(\mathbf{x}_j;\psi))\mathbf{x}_j\mathbf{x}_j^T\Big\}.$$

An estimator of (11) is then defined by plugging in estimates of unknown quantities into these expressions. Denoting these plug-in estimates by a hat leads to a variance estimator for $\hat{\boldsymbol{\beta}}_q$ of the form

$$\widehat{Var}(\hat{\boldsymbol{\beta}}_q) = n^{-1}\hat{\mathbf{B}}^{-1}(\hat{\boldsymbol{\beta}}_q)\widehat{Var}\{\Psi(\hat{\boldsymbol{\beta}}_q)\}[\hat{\mathbf{B}}^{-1}(\hat{\boldsymbol{\beta}}_q)]^T. \tag{12}$$

R routines for estimation and inference using M-quantile regression models with binary and Poisson data are available from the authors.

3·3. *M-quantile regression for a binary response: an econometric approach*

The estimating equation approach described in the previous subsection does not strictly apply to standard quantile regression for binary data, and quantile regression for binary data has been developed in the econometric literature using a latent variable concept. However, as we now show, the two approaches are very closely related, since the econometric approach can be shown to be equivalent to the solution of an estimating equation analogous to (6). Since we confine ourselves to standard quantiles in this subsection, we drop the influence function ψ from our notation and, following Kordas (2006), we assume that the observed values y_j represent the outcome of a continuously distributed latent variable. That is, the observed value y_j is generated by an unobserved (latent) real value y_j^* in the sense that $y_j = I(y_j^* > 0)$. Let $Q_q^*(\mathbf{x}_j)$ denote the conditional quantile function of this latent variable. Since $y_j = I(y_j^* > 0)$ is a monotonic transformation of y_j^*, the qth conditional quantile of y_j should be the same transformation of the qth conditional quantile of y_j^*. That is

$$Q_q(\mathbf{x}_j) = I(Q_q^*(\mathbf{x}_j) > 0).$$

Given that $Q_q^*(\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}_q$, it follows that $Q_q(\mathbf{x}_j) = I(\mathbf{x}_j^T\boldsymbol{\beta}_q > 0)$ and a 'maximum score' estimator for $\boldsymbol{\beta}_q$, defined by

$$\hat{\boldsymbol{\beta}}_q = \max_{\|\mathbf{b}=1\|} n^{-1} \sum_{j=1}^{n} \{y_j - (1-q)\} I(\mathbf{x}_j^T\mathbf{b} > 0) \tag{13}$$

was suggested by Manski (1975, 1985). Put $I_j(\mathbf{b}) = I\{y_j < I(\mathbf{x}_j^T\mathbf{b} > 0)\}$. Since $I\{y_j < I_j(\mathbf{b})\} = (1-y_j)I_j(\mathbf{b})$, we can, after some simplification, show that (13) reduces to

$$\hat{\boldsymbol{\beta}}_q = \min_{\|\mathbf{b}=1\|} n^{-1} \sum_{j=1}^{n} \Big[qI\{y_j \geqslant I_j(\mathbf{b})\} + (1-q)I\{y_j < I_j(\mathbf{b})\}\Big] |y_j - I_j(\mathbf{b})|. \tag{14}$$

This is equivalent to fitting the quantile regression model $Q_q(\mathbf{x}_j) = I(\mathbf{x}_j^T\boldsymbol{\beta}_q)$ to the observed y_j, subject to the restriction $\|\boldsymbol{\beta}_q = 1\|$, or, in what amounts to the same thing, solving (6) with $\psi(t) = sgn(t)$, subject to this restriction. Note that the restriction is necessary in order to ensure that $\boldsymbol{\beta}_q$ is identifiable (since the scale of y_j^* is unknown) and so (13) has a solution.

A smoothed version of (13) has been proposed by Horowitz (1992) as having better finite sample properties:

$$\hat{\boldsymbol{\beta}}_q = \max_{\|\mathbf{b}=1\|} n^{-1} \sum_{j=1}^{n} \{y_j - (1-q)\} F(\sigma_n^{-1}\mathbf{x}_j^T\mathbf{b}), \tag{15}$$

where F is an appropriately chosen 'smooth' cumulative distribution function defined on the entire real line and $\sigma_n \to 0$ as $n \to \infty$. The same simplifying steps as those leading to (14) allow us to write (15) as

$$\hat{\boldsymbol{\beta}}_q = \min_{\|\mathbf{b}=1\|} n^{-1} \sum_{j=1}^{n} \Big[qI\{y_j \geqslant F(\sigma_n^{-1}\mathbf{x}_j^T\mathbf{b})\} + (1-q)I\{y_j < F(\sigma_n^{-1}\mathbf{x}_j^T\mathbf{b})\}\Big] |y_j - F(\sigma_n^{-1}\mathbf{x}_j^T\mathbf{b})|,$$

since $0 < F(t) < 1 \Rightarrow I\{y_j < F(\sigma_n^{-1}\mathbf{x}_j^T\mathbf{b})\} = 1 - y_j$. That is, this 'smoothed' loss function for regression quantiles for binary data leads to essentially the same estimator as the logistic formulation (7). In fact, if we put $F(t) = \exp\{\sigma_n^{-1}\mathbf{x}_j^T\mathbf{b}\}\Big(1+\exp\{\sigma_n^{-1}\mathbf{x}_j^T\mathbf{b}\}\Big)^{-1}$ then as $\sigma_n \to 0$, $F(\sigma_n^{-1}\mathbf{x}_j^T\mathbf{b}) \to \exp\{\mathbf{x}_j^T\mathbf{b}\}\Big(1+\exp\{\mathbf{x}_j^T\mathbf{b}\}\Big)^{-1}$, and we end up with the quantile analogue of the solution to (6), with $Q_q(\mathbf{x}_j;\psi)$ defined by (7) and subject to the restriction $\|\boldsymbol{\beta}_q = 1\|$.

4. Estimation of small area proportions using M-quantile regression models

Much of the data measured in business, labour force and living conditions surveys is binary in character, and there is a growing need for reliable small area estimates based on these data. From now on therefore we focus on using M-quantile regression models for binary data to estimate the small area averages (i.e. the small area proportions) defined by a binary outcome variable.

4·1. *Point estimation*

The mixed models employed in small area estimation use random area effects to account for between-area variation. These models depend on distributional assumptions for the random part of the model and do not easily allow for outlier robust inference. An alternative approach to modelling the variability associated with the conditional distribution of Y given $\mathbf{X}$ is via M-quantile regression models. These models do not depend on strong distributional assumptions or on a predefined small area 'geography', and outlier robust inference is automatic.

A key concept when using an M-quantile regression model for small area estimation is that of the 'M-quantile coefficient' q_j for a unit $j \in s$. When the variable Y is continuous this is the value q_j such that $y_j = Q_{q_j}(\mathbf{x}_j;\psi)$. Note that M-quantile coefficients are determined at population level. If between area variation is an important part of overall population variability then units within an area will have similar M-quantile coefficients. Provided there are sample observations in area d, estimated values of their M-quantile coefficients are defined by substituting $\hat{Q}_{q_j}(\mathbf{x}_j;\psi)$ in the preceding definition. An area d specific M-quantile coefficient, $\hat{\theta}_d$ can then be defined as the average value of the estimated 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 $\bar{y}_d$ in small area d is

$$\hat{\bar{y}}_d^{MQ} = N_d^{-1}\Big\{\sum_{j\in s_d} y_{dj} + \sum_{j\in r_d}\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_{dj};\psi)\Big\}. \quad (16)$$

When Y is binary, and we model its regression M-quantile of order q via (7), the natural extension of this approach is to put $\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_{dj};\psi) = \exp\{\mathbf{x}_{dj}^T\hat{\boldsymbol{\beta}}_{\hat{\theta}_d}\}\Big(1+\exp\{\mathbf{x}_{dj}^T\hat{\boldsymbol{\beta}}_{\hat{\theta}_d}\}\Big)^{-1}$ in (16). However, this begs the question of how one defines $\hat{\theta}_d$, since the estimating equation $y_j = \hat{Q}_{q_j}(\mathbf{x}_j;\psi)$ for the estimated M-quantile coefficient of a continuous y_j no longer has a solution when y_j is binary. We therefore discuss extension of the M-quantile coefficient concept to binary Y before we consider inference based on (16).

4·2. *M-quantile coefficients for binary data*

A first step in defining M-quantile coefficients for binary data is to note that any reasonable definition of this concept has to associate a larger M-quantile coefficient with a value $y_j = 1$ compared with a value $y_j = 0$ at the same value of $\mathbf{x}_j$. The next thing to note is that the solution m_j to the equation $\hat{Q}_{m_j}(\mathbf{x}_j;\psi) = 0.5$ can be interpreted as a measure of the propensity for $y_j = 1$ to be observed relative to the propensity for $y_j = 0$ to be observed at $\mathbf{x}_j$. A value $m_j < 0.5$ indicates that $y_j = 1$ is more likely than $y_j = 0$ and vice versa. This leads to our first definition of an estimated M-quantile coefficient when Y is binary.

Definition A: Given binary data with fitted M-quantile regression function $\hat{Q}_q(\mathbf{x}_j;\psi)$, the estimated M-quantile coefficient for observation j is $q_j = (m_j + y_j)/2$, where $\hat{Q}_{m_j}(\mathbf{x}_j;\psi) = 0.5$.

Note that provided $\hat{Q}_q(\mathbf{x}_j;\psi)$ is monotone in q at $\mathbf{x}_j$, the above definition of an estimated M-quantile coefficient should be unique. In order to understand the motivation for this definition, suppose that $y_j = 0$ at $\mathbf{x}_j$ and that there are many more $Y = 0$ than $Y = 1$ 'near' $\mathbf{x}_j$. Then (a) $y_j = 0$ is not unusual, and (b) we anticipate that the monotone increasing function $f(q) = \hat{Q}_q(\mathbf{x}_j; \psi)$ will only exceed half for values of q close to one. That is, m_j will be close to one and so q_j will be slightly less than half. On the other hand, suppose $y_j = 1$ but there are still many more $Y = 0$ than $Y = 1$ 'near' $\mathbf{x}_j$. Then (a) $y_j = 1$ is unusual, and (b) we still anticipate that the monotone increasing function $f(q) = \hat{Q}_q(\mathbf{x}_j; \psi)$ will only exceed half for values of q close to one. Here q_j will be close to one. Conversely, suppose that there are many more observations with $Y = 1$ than with $Y = 0$ 'near' $\mathbf{x}_j$, so m_j is close to zero. Then if $y_j = 0$ (an unusual value) we expect q_j will also be close to zero, while if $y_j = 1$ (not unusual) we expect q_j will be slightly greater than a half.

The estimated M-quantile coefficients allow us to 'index' the sample data. A somewhat different indexing based on quantile regression modelling of Y is described in Kordas (2006). This takes a latent variable approach and the resulting index is essentially defined by a quantile-based estimate of $P(y_j = 1|\mathbf{x}_j)$. Under linearity of the conditional quantiles of this latent variable, we have already seen that $Q_q(\mathbf{x}_j) = I(\mathbf{x}_j^T \boldsymbol{\beta}_q > 0)$ and so $P(y_j = 1|\mathbf{x}_j) = 1 - h_j$, where $\mathbf{x}_j^T \boldsymbol{\beta}_{h_j} = 0$. Consequently, given an estimate $\hat{\boldsymbol{\beta}}_q$ for each value $0 < q < 1$ we can index the sample observations by $p_j = 1 - h_j$ where $\mathbf{x}_j^T \hat{\boldsymbol{\beta}}_{h_j} = 0$. Note that this index does not depend on y_j, and so cannot reflect individual effects, which would seem to limit its usefulness in characterising how groups differ after covariate effects have been taken into account. However, we can use the approach leading to Definition A to extend this index by allowing it to reflect individual effects.This leads to our second definition of an estimated M-quantile coefficient for the binary case.

DEFINITION B: Given binary data with fitted M-quantile regression function $\hat{Q}_q(\mathbf{x}_j; \psi)$, the estimated M-quantile coefficient for observation j is $q_j = (h_j + y_j)/2$, where $\mathbf{x}_j^T \hat{\boldsymbol{\beta}}_{h_j} = 0$.

Note that if $\mathbf{x}_j^T \hat{\boldsymbol{\beta}}_q = 0 \Leftrightarrow \hat{Q}_q(\mathbf{x}_j; \psi) = 0.5$ then Definition B and Definition A are identical. This condition will hold, for example, whenever ψ is the identity function and $Q_q(\mathbf{x}_j; \psi) = Q_q(\mathbf{x}_j) = F(\mathbf{x}_j^T \boldsymbol{\beta}_q)$ where $F(t)$ is a distribution function that satisfies $F(0) = 0.5$.

Unfortunately, both Definition A and Definition B have a serious deficiency. This follows from the fact that in applications where h_j varies around some constant, say h, q_j will be 'concentrated' near $(1 + h)/2$ and $h/2$. Furthermore, it is impossible to observe $q_j = 0.5$ in general. An extreme case is where there is no relationship between y_j and $\mathbf{x}_j$, and $y_j = 1$ is just as likely as $y_j = 0$. In this case $h_j = 0.5$, and there are just two possible values of q_j, 0.75 $(y_j = 1)$ and 0.25 $(y_j = 0)$.

The basic reason for this behaviour is that both Definition A and Definition B compute q_j on the same scale as y_j. This makes sense when the distribution of y_j is measured on a linear scale. However, in the binary case the distribution of y_j is linear in the logistic scale, and so it makes sense to define q_j in the same way. That is, we replace q_j and h_j in Definition B by $\hat{Q}_{q_j}(\mathbf{x}_j; \psi)$ and $\hat{Q}_{0.5}(\mathbf{x}_j; \psi)$ respectively, leading to our third, and final, definition of q_j:

DEFINITION C: Given binary data with fitted M-quantile regression function $\hat{Q}_q(\mathbf{x}_j; \psi)$, the estimated M-quantile coefficient for observation j is q_j, where $\hat{Q}_{q_j}(\mathbf{x}_j; \psi) = (\hat{Q}_{0.5}(\mathbf{x}_j; \psi) + y_j)/2$.

Note that under a logistic specification for $\hat{Q}_q(\mathbf{x}_j; \psi)$, using Definition C is equivalent to defining q_j as the solution to $y_j^* = \mathbf{x}_j^T \boldsymbol{\beta}_{q_j}$, where

$$y_j^* = \log\Big(\frac{0.5\{\hat{Q}_{0.5}(\mathbf{x}_j; \psi) + y_j\}}{1 - 0.5\{\hat{Q}_{0.5}(\mathbf{x}_j; \psi) + y_j\}}\Big).$$

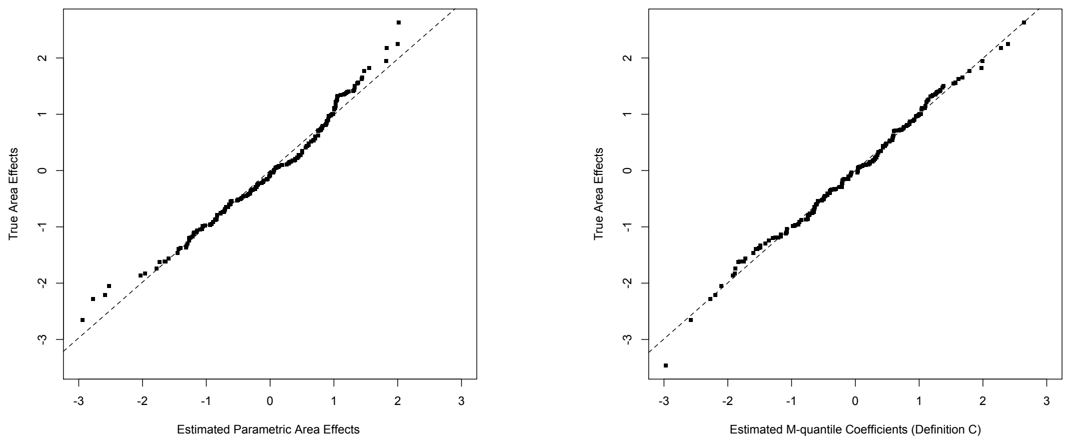

Fig. 1. Estimated area effects vs. true area effects (left plot) and estimated M-quantile coefficients vs. true area effects (right plot) in a Monte-Carlo simulation with $D = 200$ and $n_d = 25$.

The value y_j^* above can be thought of as a pseudo-value that behaves ‘like’ the unobservable latent variable whose distribution determines that of y_j. In the rest of this paper, and particularly in the simulation experiments reported in Section 5, we use Definition C when calculating estimated M-quantile coefficients.

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 binary M-quantile model proposed in this paper. A natural question is then the strength of the relationship between the actual area effects and the estimated M-quantile coefficients. Some empirical evidence for such a relationship is displayed in Figure 1. These scatterplots show how area effects estimated using the `glmer` function in R and M-quantile coefficients estimated via Definition C are related to true area effects. The simulated data underpinning these plots were generated using $D = 200$ areas, each with a sample size of $n_d = 25$. At each simulation, values of x_{dj} were independently drawn as $Normal(0, 1)$ and corresponding values of y_{dj} were then generated as $Bernoulli(p_{dj})$ with $p_{dj} = \exp\{\eta_{dj}\}(1 + \exp\{\eta_{dj}\})^{-1}$ and $\eta_{dj} = x_{dj} + u_d$. The small area effects u_d were independently drawn as $Normal(0, 1)$. Figure 1 shows how the estimated area effects and the estimated M-quantile coefficients are related to the true area effects in one Monte-Carlo simulation. Over $1,000$ simulations, the average correlation between the true area effects and estimated area effects was 0.89, and the corresponding average correlation between the true area effects and the estimated M-quantile coefficients was 0.80. These results suggest 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. Note also that these simulations build on data generated via a GLMM. In real applications, where GLMM assumptions may be violated, we expect that an M-quantile approach should offer a robust alternative for small area estimation.

4·3. *Mean squared error estimation*

To start, we propose a MSE estimator for (16) based on the linearisation approach set out in Chambers et al. (2009). This assumes that the working model for inference conditions on the realised values of the area effects, and so the MSE of interest is conditional and equal to a conditional prediction variance plus a squared conditional prediction bias. In order to conserve space, we omit some technical details in the following development, but these are available from the authors upon request. We also assume that the estimated area-level M-quantile coefficient values θ_d have negligible variability and so can be treated as fixed. A first order approximation to the conditional prediction variance of (16) is then

$$\begin{aligned} Var(\hat{\bar{y}}_d^{MQ} - \bar{y}_d|\theta_d) &= N_d^{-2}\Big\{Var\Big[\sum_{j\in r_d}\hat{Q}_{\theta_d}(\mathbf{x}_j;\psi)\Big] + \sum_{j\in r_d} Var(y_j)\Big\} \\ &\approx N_d^{-2}\Big\{\Big[\sum_{j\in r_d} Q_{\theta_d}(\mathbf{x}_j;\psi)\mathbf{x}_j^T\Big] Var(\hat{\boldsymbol{\beta}}_{\theta_d})\Big[\sum_{j\in r_d} Q_{\theta_d}(\mathbf{x}_j;\psi)\mathbf{x}_j^T\Big]^T \\ &\quad + \sum_{j\in r_d} Var(y_j)\Big\}, \end{aligned}$$

which can be estimated by

$$\begin{aligned} \widehat{Var}(\hat{\bar{y}}_d^{MQ}) &= N_d^{-2}\Big\{\Big[\sum_{j\in r_d}\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_j;\psi)\mathbf{x}_j^T\Big]\widehat{Var}(\hat{\boldsymbol{\beta}}_{\hat{\theta}_d})\Big[\sum_{j\in r_d}\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_j;\psi)\mathbf{x}_j^T\Big]^T \\ &\quad + \sum_{j\in r_d}\widehat{Var}(y_j)\Big\}. \end{aligned}$$

Here $\widehat{Var}(\hat{\boldsymbol{\beta}}_{\hat{\theta}_d})$ is the sandwich-type estimator (12) and $\widehat{Var}(y_j)$ can be calculated either by (i) using the sample data from area d, $\widehat{Var}(y_j) = \hat{\bar{y}}_d(1-\hat{\bar{y}}_d)$ or by (ii) pooling data from the entire sample, in which case $\widehat{Var}(y_j) = \hat{\bar{y}}(1-\hat{\bar{y}})$. Note that the pooled estimator should lead to more stable prediction variance estimates when area sample sizes are very small.

The conditional prediction bias can be approximated using the results of Copas (1988):

$$\begin{aligned} E(\hat{\bar{y}}_d^{MQ} - \bar{y}_d|\theta_d) &\approx -\frac{1}{2N}\Big\{\frac{\partial}{\partial\boldsymbol{\beta}_{\theta_d}}\Psi(\boldsymbol{\beta}_{\theta_d})\Big\}^{-1}\Big\{tr\Big[\Big\{\frac{\partial}{\partial\boldsymbol{\beta}_{\theta_d}\partial\boldsymbol{\beta}_{\theta_d}^T}\Psi(\boldsymbol{\beta}_{\theta_d})\Big\}Var(\hat{\boldsymbol{\beta}}_{\theta_d})\Big]\Big\} \\ &\quad \Big\{\frac{\partial}{\partial\boldsymbol{\beta}_{\theta_d}}\sum_{j\in r_d} Q_{\theta_d}(\mathbf{x}_j;\psi)\Big\}, \end{aligned}$$

with corresponding plug-in estimator

$$\begin{aligned} \widehat{Bias}(\hat{\bar{y}}_d^{MQ}) &= -\frac{1}{2N}\Big\{\frac{\partial}{\partial\boldsymbol{\beta}_{\theta_d}}\Psi(\boldsymbol{\beta}_{\theta_d})|_{\boldsymbol{\beta}_{\theta_d}=\hat{\boldsymbol{\beta}}_{\hat{\theta}_d}}\Big\}^{-1}\Big\{tr\Big[\Big\{\frac{\partial}{\partial\boldsymbol{\beta}_{\theta_d}\partial\boldsymbol{\beta}_{\theta_d}^T}\Psi(\boldsymbol{\beta}_{\theta_d})|_{\boldsymbol{\beta}_{\theta_d}=\hat{\boldsymbol{\beta}}_{\hat{\theta}_d}}\Big\}\widehat{Var}(\hat{\boldsymbol{\beta}}_{\hat{\theta}_d})\Big]\Big\} \\ &\quad \Big\{\frac{\partial}{\partial\boldsymbol{\beta}_{\theta_d}}\sum_{j\in r_d} Q_{\theta_d}(\mathbf{x}_j;\psi)|_{\boldsymbol{\beta}_{\theta_d}=\hat{\boldsymbol{\beta}}_{\hat{\theta}_d}}\Big\}. \end{aligned}$$

The estimator of the conditional MSE of $\hat{\bar{y}}_d^{MQ}$ is then

$$mse^A(\hat{\bar{y}}_d^{MQ}) = \widehat{Var}(\hat{\bar{y}}_d^{MQ}) + \{\widehat{Bias}(\hat{\bar{y}}_d^{MQ})\}^2. \quad (17)$$

We also propose two different bootstrap-based methods for estimating the MSE of (16) - a nonparametric bootstrap and a random effect block bootstrap. In order to save space, the computational details of each bootstrap procedure are set out in the Appendix. Here we summarise the main characteristics of each method.

The nonparametric bootstrap-based estimator (hereafter NPB) of the MSE of $\hat{\bar{y}}_d^{MQ}$ is defined by using the GLMM representation (2) to generate bootstrap populations of binary values that mimic the real population, sampling from this bootstrap population and calculating (16). This procedure is an extension to binary data of the method proposed by Tzavidis et al. (2010). The bootstrap procedure is nonparametric in the sense that it replicates area effects in the population by expressing the linear component of the logistic M-quantile regression model (7) as

$$\gamma_{jq} = \mathbf{x}_j^T \boldsymbol{\beta}_{0.5} + \bar{\mathbf{x}}_d^T (\boldsymbol{\beta}_{\theta_d} - \boldsymbol{\beta}_{0.5}), \tag{18}$$

where $\bar{\mathbf{x}}_d^T$ is the vector of area d averages of the model covariates. The last term on the right-hand side of (18) is then interpreted as a pseudo-random effect for area d.

The random effect block bootstrap, see Chambers & Chandra (2012), is a robust alternative to parametric bootstrap methods for clustered data. It is free of both the distribution and the dependence assumptions of the usual parametric bootstrap for such data and is consistent when the mixed model assumption is valid. In particular, it preserves area effects by bootstrap resampling within areas. We adapt this procedure (hereafter REBB) for estimating the distribution of the M-quantile predictor (16) by resampling the marginal logistic scale residuals $r_{dj}^{MQ} = \mathbf{x}_{dj}^T(\hat{\boldsymbol{\beta}}_{\hat{\theta}_d} - \hat{\boldsymbol{\beta}}_{0.5})$ within each area to generate bootstrap values of $P(y_j = 1|\mathbf{x}_j)$ for the population units making up the area. Bootstrap binary population values are then obtained using Bernoulli simulation.

5. Simulation studies

We present results from two types of simulation studies that are used to examine the performance of the small area estimators discussed in the preceding Sections. In Subsection 5·1 below we report results from model-based simulations where population data are first generated using a GLMM and a sample is then drawn from this simulated population using a pre-specified design. The properties of estimators of small area proportions are assessed by comparing the estimates that they generate based on the sample data to the corresponding population proportions. In Subsection 5·2 we report results from a design-based simulation. Here a realistic finite population is simulated by resampling from data collected in a sample survey, and then repeated samples are drawn from it using the same (or similar) sample design as the original survey. In this case the survey data were sourced from the European Union Statistics on Income and Living Conditions (EU-SILC) 2005 survey, and the distribution of estimates of small area proportions generated under repeated sampling is used to compare the properties of different estimators of these proportions.

Two different M-quantile versions of (16) were investigated in the simulations, both based on a linear logistic M-quantile model defined by a Huber influence function with tuning constant c. In the first, referred to as M-quantile below, $c = 1.345$, while the second, referred to as Expectile below, $c = 100$. These estimators were compared with the EBP (3) under a GLMM with logistic link function and with the direct estimator (the sample proportion). Both MSE estimation and confidence interval coverage performance were evaluated using the analytic and two bootstrap methods described in Subsection 4·3 for the M-quantile predictor.

Note that the logistic M-quantile linear regression fit underpinning the M-quantile and Expectile predictors was obtained using an extended version of a M-quantile linear regression model function for SAE written in *R*. The parameters of the GLMM used in the EBP were estimated using the function `glmer` in *R*.

5·1. *Model-based simulations*

In each simulation we generated $N = 5,000$ population values of X and Y in $D = 50$ small areas with $N_d = 100,\ d = 1,\dots,D$. Individual x_{dj} values were drawn independently at each simulation as $Uniform(a_d, b_d)$, for $a_d = -1$ and $b_d = d/4$, $d = 1,\dots,D,\ j = 1,\dots,N_d$. Values of y_{dj} were then generated as $Bernoulli(p_{dj})$ with $p_{dj} = \exp\{\eta_{dj}\}(1+\exp\{\eta_{dj}\})^{-1}$ and $\eta_{dj} = x_{dj}\beta + u_d$. The small area effects u_d were independently drawn from a normal distribution with mean 0 and variance $\varphi = 0.25$, and $\beta = 1$ (González-Manteiga et al., 2007). Population values generated under this scenario are denoted by (0) below. In addition, we generated data corresponding to a combined misclassification error and measurement error scenario, denoted (M) below. In this scenario, a random 1% sample of the x_{dj} values were replaced by 20 (introducing measurement error) and the corresponding y_{dj} values were set to 0 (introducing misclassification error).

For each of these scenarios $T = 1,000$ Monte-Carlo populations were generated. For each generated population and for each area d we then took simple random samples without replacement of sizes $n_d = 10$ and $n_d = 20$ so that the overall sample sizes were $n = 500$ and $n = 1,000$. For each sample the M-quantile and Expectile predictors, the EBP and the direct estimator were used to estimate the small area proportions $\bar{y}_d,\ d = 1,\dots,D$.

The performances of different small area estimators for area d were evaluated with respect to two criteria: their average error $T^{-1}\sum_{t=1}^{T}(\hat{\bar{y}}_{dt} - \bar{y}_{dt})$ and the square root of their average squared error $T^{-1}\sum_{t=1}^{T}(\hat{\bar{y}}_{dt} - \bar{y}_{dt})^2$. These are denoted Bias and RMSE respectively below. Here $\bar{y}_{dt}$ denotes the actual area d value at simulation t, with predicted value $\hat{\bar{y}}_{dt}$. The median values of Bias and RMSE over the D small areas are set out in Table 1, where we see that claims in the literature (Chambers & Tzavidis, 2006) about the superior outlier robustness of the M-quantile predictor compared with the EBP and the Expectile predictor certainly hold true in these simulations. In particular, under the (0) scenario the EBP performs better than the M-quantile and Expectile predictors in terms of Bias, whereas the M-quantile predictor is the best under the (M) scenario. In terms of RMSE, there is nothing to choose between EBP, M-quantile and Expectile under the (0) scenario, while under the (M) scenario the M-quantile predictor is clearly superior.

Table 1. *Model-based simulation results: Predictors of small area proportions.*

	$n_d = 10$		$n_d = 20$	
Predictor/Scenario	(0)	(M)	(0)	(M)
	Median values of Bias			
EBP	0.0013	-0.0200	0.0008	-0.0116
M-quantile	0.0041	0.0046	0.0041	0.0045
Expectile	0.0043	-0.0178	0.0045	-0.0164
Direct	0.0004	-0.0001	0.0001	-0.0001
	Median values of RMSE			
EBP	0.0519	0.0598	0.0442	0.0507
M-quantile	0.0509	0.0511	0.0444	0.0445
Expectile	0.0506	0.0625	0.0442	0.0508
Direct	0.1146	0.1148	0.0770	0.0777

In order to evaluate the performances of the MSE estimators proposed in Subsection 4·3 we used the data generated for the scenario with $D = 50$ and $n_d = 10$ and also carried out a further model-based simulation study with the same sample sizes within the small areas but with $D = 100$ and $N_d = 100$. Again, $T = 1,000$ Monte-Carlo populations were generated, with individual x_{dj} values drawn independently as $Uniform(a_d, b_d)$, with $a_d = -1$ and $b_d = d/8$, $d = 1, \ldots, D,\ j = 1, \ldots, N_d$. For each generated population a simple random sample without replacement of size $n_d = 10$ was drawn from each area d, the M-quantile predictor calculated as well as its linearisation MSE estimator (17) and the two bootstrap MSE estimators NPB and REBB, both based on 100 bootstrap iterations. The behaviour of these MSE estimators for each scenario is displayed in Table 2 where we show the medians of their area specific Bias and RMSE values, expressed in relative terms (%). We also show the median empirical coverage rates for nominal 95 per cent confidence intervals based on these methods. In the case of (17) these intervals were defined by the small area estimate plus or minus twice the value of the square root of (17). For NPB and REBB these intervals were based on the 2.5 and the 97.5 percentiles of the relevant bootstrap distribution.

Examination of Table 2 shows that all three MSE estimation methods tend to be biased low, but all generate nominal 95 per cent confidence intervals with acceptable coverage. Overall, the REBB bootstrap estimator seems preferable because it shows smaller or similar bias and more stability than both the linearisation-based estimator (17) and the NPB bootstrap estimator.

Table 2. *Model-based simulation results: MSE estimators.*

	Median values of Relative Bias (%)		*Median values Relative RMSE (%)*		*Median values of Coverage Rate (%)*	
			$D = 50$			
Estimator/Scenario	(0)	(M)	(0)	(M)	(0)	(M)
$mse^{A}(\hat{\bar{y}}_d^{MQ})$	-5.37	-6.05	24.33	25.30	95	94
$mse^{NPB}(\hat{\bar{y}}_d^{MQ})$	-12.06	-11.91	15.64	16.40	92	92
$mse^{REBB}(\hat{\bar{y}}_d^{MQ})$	-7.57	-1.03	12.70	13.00	93	95
			$D = 100$			
$mse^{A}(\hat{\bar{y}}_d^{MQ})$	-5.80	-6.91	25.15	24.51	95	94
$mse^{NPB}(\hat{\bar{y}}_d^{MQ})$	-11.42	-11.28	15.10	14.72	92	93
$mse^{REBB}(\hat{\bar{y}}_d^{MQ})$	-6.56	-0.86	12.02	11.49	94	95

5·2. *Design-based simulation*

The population underpinning the design-based simulation is based on data collected from a sample of $1,560$ households spread across 64 municipalities (out of 287) in Tuscany They were collected by ISTAT as part of the European Union Statistics on Income and Living Conditions (EU-SILC) 2005 survey. These survey data provide information on a variety of issues related to living conditions of the people in Tuscany, including details on income and non-income dimensions of poverty in the region, and form the basis of poverty assessment in this region. Here we define the sampled municipalities as our small areas of interest. Since nine municipalities had sample sizes less than three, these were combined with adjacent municipalities, leading to a total of 55 small areas.

We used these sample households to generate a population of $N = 74,951$ households by sampling with replacement from the original sample of $1,560$ households with probabilities proportional to their sample weights. Given this (fixed) population, we then independently drew $1,000$ stratified random samples from it, with the 55 (redefined) municipalities serving as strata, and with the sample size for each municipality fixed to be the same as in the original sample. Note that these sample sizes varied from 4 to 116. The binary variable of interest Y was defined

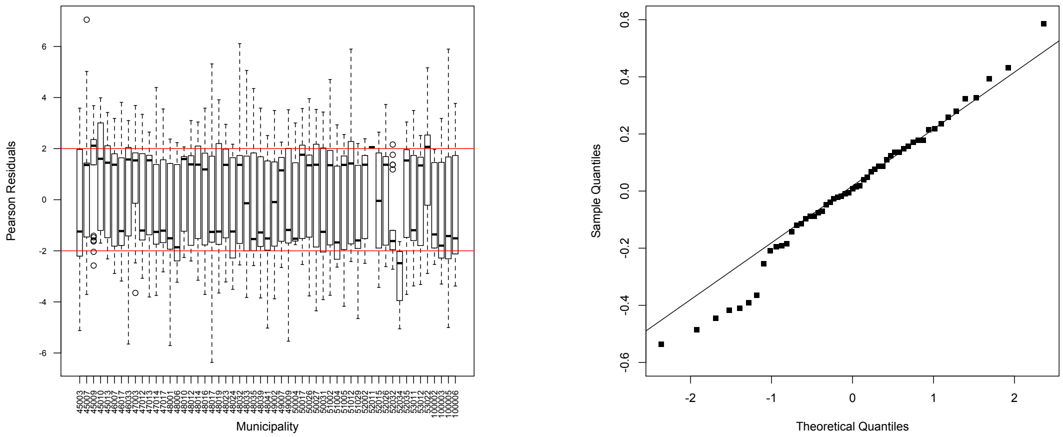

Fig. 2. Model fit diagnostics for a logistic mixed model fit to the EU-SILC data: This shows the distribution of Pearson residuals by municipality (left) and a normal probability plot of the estimated municipality effects (right).

as taking the value 1 if the equivalised income of a household was below the median income for the simulated population and 0 otherwise. The aim was to compare the repeated sampling performance of different predictors of the proportion of households below median equivalised income at municipality level, using household ownership status (a strong indicator of poverty), the age of the head of the household, the employment status of the head of the household, the gender of the head of the household, the years of education of the head of the household and the household size as covariates.

Figure 2 shows selected diagnostics for a logistic mixed model fit to Y based on the original EU-SILC data and using these covariates. The distribution of Pearson residuals indicates the presence of potential influential observations in the data, with a number of large residuals ($|r_{dj}| > 2$). Further evidence for the presence of these influential observations is obtained when we fit the model using a robust method (Cantoni & Ronchetti, 2001) and note that although most observations receive a weight of 1 in this fit, there are 19 (about 1% of the overall sample) that receive a weight of less than 0.7. The normal probability plot shown in Figure 2 also indicates that the Gaussian assumption for the distribution of the random effects in this GLMM is doubtful. Using a model that relaxes this assumption, such as an M-quantile model with a bounded influence function, therefore seems reasonable for these data.

Table 3 shows the five point summaries of the municipality level distributions of Bias and RMSE generated under repeated sampling for the same four estimators that were the focus of the model-based simulations reported in the previous Subsection. These show that the M-quantile and Expectile predictors have similar performance and both generally perform better than the EBP and the Direct estimator in terms of RMSE. In particular, if one focuses on median performance, then the M-quantile predictor seems the best of the four.

Table 4 shows the median values of Bias and RMSE, expressed in relative terms, of the three MSE estimators for the M-quantile predictor. In this case the NPB method performs best, with the REBB and the linearisation-based estimator (17) displaying a somewhat larger positive Bias

Table 3. *Design-based simulation results: Distributions of Bias and RMSE for predictors of the proportion of households with below median income.*

Predictor	Min	Q1	Median	Mean	Q3	Max
			Bias			
EBP	-0.4692	-0.0230	0.0026	-0.0051	0.0187	0.3587
M-quantile	-0.1882	-0.0359	0.0007	-0.0007	0.0384	0.1991
Expectile	-0.1851	-0.0317	0.0031	-0.0012	0.0364	0.2227
Direct	-0.0049	-0.0010	0.0004	0.0003	0.0016	0.0068
			RMSE			
EBP	0.0401	0.0631	0.0703	0.0870	0.0806	0.4701
M-quantile	0.0264	0.0566	0.0647	0.0801	0.0965	0.2041
Expectile	0.0265	0.0555	0.0654	0.0807	0.0964	0.2316
Direct	0.0011	0.0816	0.0955	0.0911	0.1046	0.1507

and slightly more instability. Again, we see that median coverage performance of all three MSE estimation methods appears acceptable.

Table 4. *Design-based simulation results: Performances of MSE estimators.*

Estimator	*Median Relative Bias (%)*	*Median Relative RMSE (%)*	*Median Coverage Rate (%)*
$mse^{A}(\hat{\bar{y}}_d^{MQ})$	11.03	30.32	93
$mse^{NPB}(\hat{\bar{y}}_d^{MQ})$	6.72	27.70	97
$mse^{REBB}(\hat{\bar{y}}_d^{MQ})$	11.28	28.46	97

6. Applications

6·1. *Estimates of ILO unemployment for UALADs of Great Britain*

In this Subsection we apply the M-quantile modelling approach to estimating the number of unemployed people aged 16 and over in each of 406 Unitary Authorities and Local Authority Districts (UALADs) of Great Britain. We use the ILO unemployment definition and data from a sample of about $169,000$ people aged 16 and over who participated in the UK Labour Force Survey (LFS) carried out by the Office for National Statistics (ONS) in 2000, with M-quantile model covariates specified by prior studies of small area estimation of UK labour force characteristics (Molina et al., 2007). In particular, we use the following covariates: sex-age category of an individual (6 categories corresponding to Male/Female and age groups $16-25, 26-40$ and > 40), government office region of the UALAD (12 categories), ONS socio-economic classification of the UALAD (7 categories) and total of registered unemployed in the sex-age group for the UALAD. In order to compare M-quantile estimates with EBP estimates, the same covariates were also used to fit a logistic mixed model to the LFS data, and corresponding EBP-type estimates were computed. All estimates of proportions were scaled up to estimates of counts by multiplying by population counts for each sex-age group in a UALAD and then summing over these groups within a UALAD.

Figure 3 shows the normal probability plot of estimated UALAD random effects obtained by fitting a logistic mixed model to the sample data. A Shapiro-Wilk normality test rejects the null hypothesis of normality for these estimated random effects (p-value $= 0.01371$). Applying a robust fitting method (Cantoni & Ronchetti, 2001) to this data set, we note that although most observations receive a weight of 1, about 3.5% receive weights less than 0.7. Using a less parametric model, e.g. an M-quantile model with a bounded influence function, therefore seems reasonable for these data.

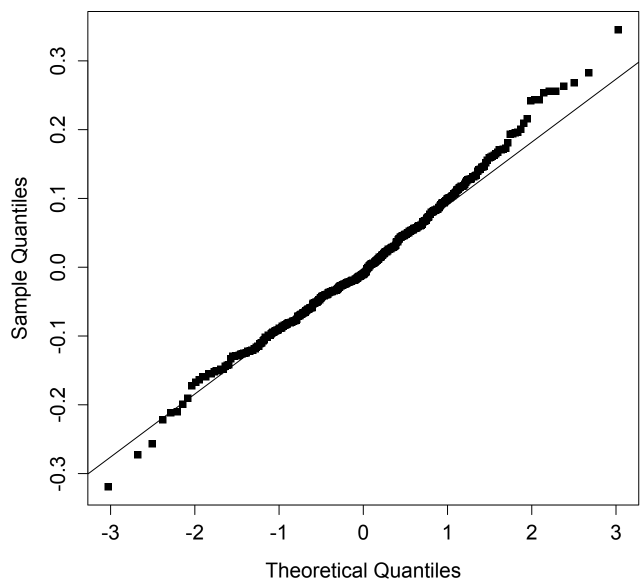

Fig. 3. Normal probability plot of estimated UALAD random effects for proportion of unemployed population aged 16 and over, based on a logistic mixed model fit to UK LFS data.

In order to assess the resulting M-quantile estimates we note that model-based small area estimates should be both consistent with corresponding unbiased direct estimates and more precise than them. Figure 4 plots the M-quantile estimates of total numbers of unemployed against corresponding direct estimates for each UALAD. We can see that the M-quantile estimates appear to be generally consistent with the direct estimates, although the relationship between these two sets of estimates diverges in the larger UALADs, where the M-quantile estimates tend to be larger. This is consistent with the right skewed distribution of estimated area effects shown in Figure 3. Overall, however, the correlation of 0.78 between the M-quantile estimates and the direct estimates is reasonably high.

In order to assess the gain in precision from using model-based estimates instead of the direct estimates, we can look at the distribution of the ratios of the estimated coefficients of variation (CVs) of the direct and the model-based estimates. A value greater than 1 for this ratio indicates that the estimated CV of the model-based estimate is less than that of the direct estimate. Figure 5 shows the relationship between these ratios and the number of unemployed people in the LFS sample in each UALAD. Two sets of ratios are plotted - those corresponding to EBP estimates (red) and those corresponding to M-quantile estimates (blue). Note that the CV for the M-quantile estimate is calculated as $[mse^{REBB}(\hat{\hat{y}}_d^{MQ})]^{1/2}/\hat{\hat{y}}_d^{MQ}$, while that for the EBP esti-

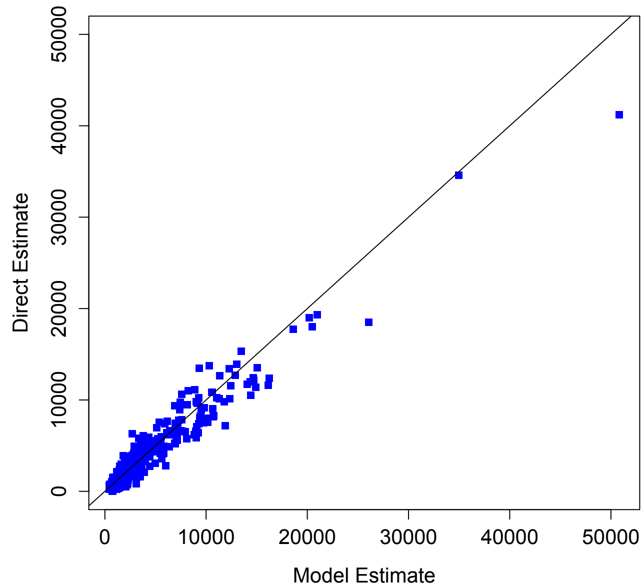

Fig. 4. Numbers of unemployed people aged 16 and over in UALADs of Great Britain in 2000: Model-based M-quantile estimates versus corresponding direct estimates.

mate is calculated as $[mse^{EBP}(\hat{\bar{y}}_d^{EBP})]^{1/2}/\hat{\bar{y}}_d^{EBP}$, where $mse^{EBP}(\hat{\bar{y}}_d^{EBP})$ is obtained using the bootstrap procedure described in González-Manteiga et al. (2007).

It is clear from Figure 5 that the estimated CVs of the M-quantile and EBP estimates of unemployment are generally much lower than those of the direct estimates. Furthermore, the estimated CVs of the M-quantile estimates are generally lower than those of the EBP estimates. This may be due to of the small number of sampled unemployed individuals within the UALADs and consequent problems with estimation of the area random effects when fitting the logistic mixed model using LFS data. Overall, our conclusion is that for these data, and for estimating UALAD unemployment, the M-quantile approach seems superior to the EBP approach.

We also carried out a similar exercise where we used these UK LFS data to estimate the total number of employed people in each UALAD. In this case the much larger samples of employed people lead to the conclusion that there was no significant advantage in using model-based methods (either EBP or M-quantile) compared with direct estimation. We do not provide further details here, but they are available from the authors on request.

6·2. *Estimates of number of households in poverty in Local Labour Systems of Tuscany*

In this Subsection we describe a second application of the M-quantile modelling approach to small area estimation. In this case we focus on estimating the number of poor households in each of the 57 Local Labour Systems (LLS's) of Tuscany, using the 2005 EU-SILC dataset described in Section 5·2. Poverty maps based on such measures are important tools for providing information on the spatial distribution of poverty, and are often used to assist the implementation of poverty alleviation programs. In this application a household is defined to be poor if its equivalised income falls below a minimum level (the poverty line) necessary to meet basic food and

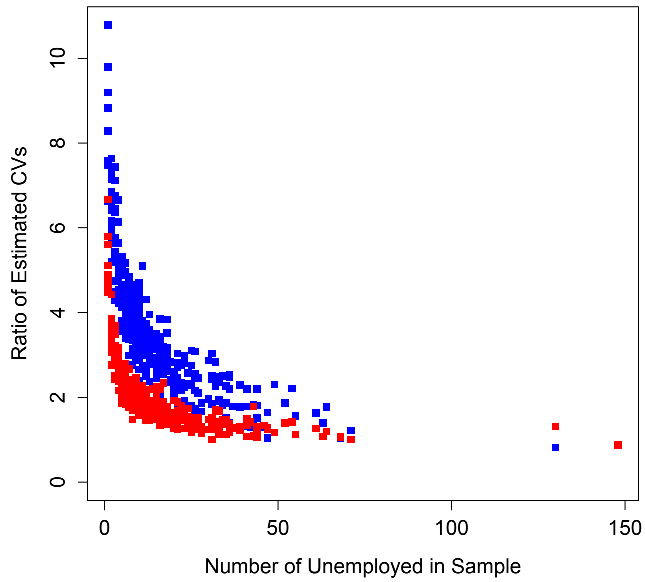

Fig. 5. Ratio of estimated coefficients of variation of direct estimates to M-quantile (blue) and EBP (red) estimates of total number of unemployed for each UALAD.

non-food needs. The poverty line is defined as 60% of median equivalised income. There are 57 LLSs in total in Tuscany, with 28 not represented in the EU-SILC dataset. Direct estimates of the number of poor households within each LLS level have high variances, particularly for LLS's with small sample sizes. Moreover, direct estimates cannot be provided for areas with no sample.

Estimated numbers of poor households for each LLS were calculated using the M-quantile predictor (16) with the logistic M-quantile specification (7) as well as the EBP predictor (3) based on a logistic mixed model, with covariates defined by ownership status, age of the head of the household, gender of the head of the household and their interactions. Population data for these covariates were drawn from the Population Census 2001. An issue with this approach is the potential lack of comparability between household-level variables measured in the 2001 Population Census and the same variables measured in the the 2005 EU-SILC. However, the covariates used in this study are not expected to change significantly over a short period of time.

Figure 6 shows the normal probability plot of the estimated LLS random effects obtained from the logistic mixed model fit to the EU-SILC data. It indicates that their distribution is left-skewed and that the assumption of normality is probably incorrect. As already noted in 5·2, there is evidence of outlying data values in this dataset. Consequently, we fit an M-quantile model defined by a Huber Proposal 2 influence function ($c = 1.345$) to the EU-SILC data in order to estimate LLS counts of number of poor households.

The correlation between the direct and M-quantile estimates for the sampled LLS's is 0.97, while the corresponding correlation for the EBP estimates is 0.95. As Figure 7 shows, the overall consistency between the M-quantile estimates and the direct estimates is very good, although this tends to become weaker for LLS's with large estimated number of poor households.

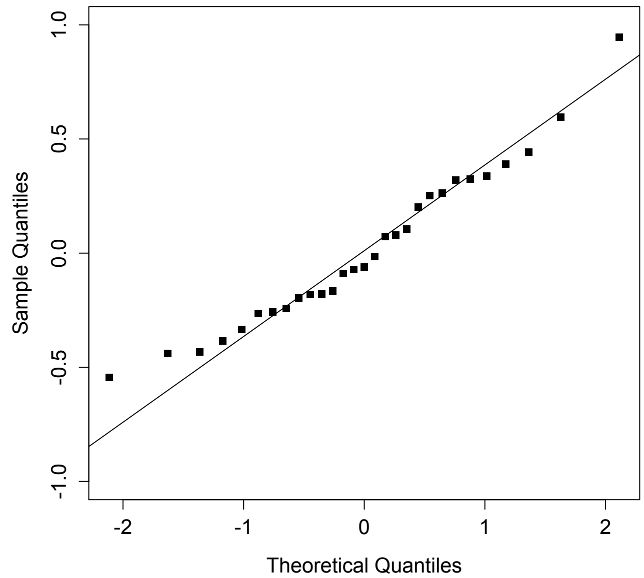

Fig. 6. Normal probability plot of estimated LLS-level residuals from logistic mixed model fit to EU-SILC data

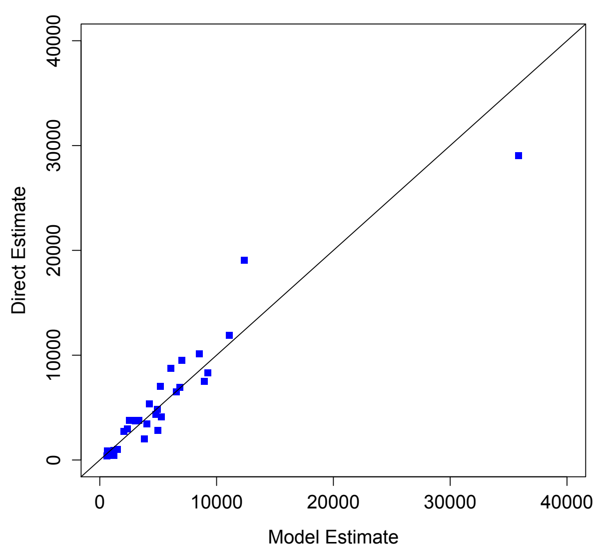

Fig. 7. Model-based M-quantile estimates of numbers of households in poverty in sampled LLS's compared with corresponding direct estimates.

The left plot in Figure 8 displays the values of the ratios of the estimated CVs for the direct and model-based estimates of numbers of poor households and the right plot shows the distribution

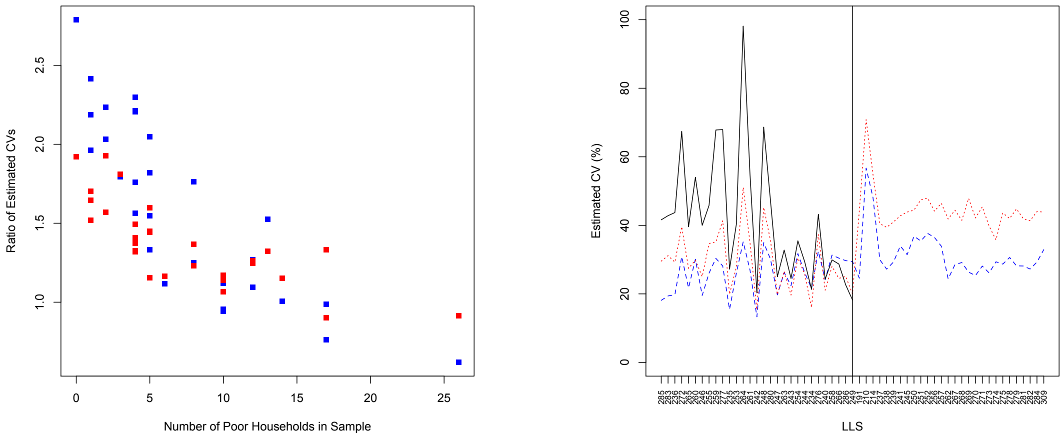

Fig. 8. Left plot shows ratio of estimated coefficients of variation of direct estimates to M-quantile (blue) and EBP (red) estimates of total number of poor households for each LLS and right plot shows distribution of LLS values of estimated CV for direct (solid line) and model-based estimates, with estimated CVs for the M-quantile predictor shown as a dashed blue line and estimated CVs for the EBP shown as a dashed red line.

across LLS's of the estimated CVs (expressed in percentage terms) of these estimates. In both plots blue indicates M-quantile estimates and red indicates EBP estimates. The solid black line in the right plot refers to the direct estimates. The estimated gains of the model-based estimates over the direct estimates is not as striking as in the previous application, but are still substantial, particularly for LLS's with a small number of sampled households. Generally, the M-quantile estimates have a smaller estimated CV than corresponding EBP estimates, with the most striking difference being for the non-sampled LLS's (located to the right of the vertical dividing line), where the M-quantile estimates are very stable and clearly dominate the EBP estimates.

7. Final remarks

Models for discrete data, and specifically for binary data, are important in small area estimation. In this paper we extend the concept of M-quantile regression modelling to discrete data, and show how these models can be used in small area estimation. By construction, this approach is outlier robust, does not require strong assumptions about the distribution of the response and can be applied independently of any pre-specified grouping of the data. In particular, the same fitted M-quantile model can be used to provide estimates for more than one small area 'geography' defined on a target population.

In a number of model-based and design-based simulations, as well as in two realistic applications, we compare the performance of small area estimates of proportions based on a logistic M-quantile specification with those defined by a plug-in approach to empirical best prediction based on a logistic mixed model specification. Our results indicate that the M-quantile estimates are either comparable or preferable. We also introduce three different approaches to estimating the mean squared error of the M-quantile model-based estimator, one based on a linearisation argument and the other two based on bootstrap. All three MSE estimators provide acceptable coverage performance in our simulations, with the second bootstrap method (REBB), based on the application of a modified block bootstrap, being slightly preferable because of its stability and simplicity.

An obvious extension of the development set out in this paper is to M-quantile versions of GLMs for count data. Although we briefly describe the M-quantile extension to GLM-type modelling for a Poisson variable, we do not explore the behaviour of corresponding small area estimators for count data based on the M-quantile approach. This is an area of current research. We also do not explore the extension of M-quantile modelling to multi-category data (e.g. multinomial data). This remains an open problem.

APPENDIX

Let A denote a set of objects and let m denote a strictly positive integer. In what follows, we use the notation $srswr(A, m)$ to denote the set of size m obtained by sampling with replacement m times from the set A.

Nonparametric bootstrap (NPB) procedure

Given a finite population U with values y_{dj} (a binary variable) and a sample s drawn from it, the steps of the NPB procedure are as follows:

- (Step 1) From sample s, fit model (7) to the initial data and obtain predictors $\hat{\bar{y}}_d^{MQ}$. For each small area compute the pseudo-random effect $\hat{u}_d^{MQ} = \bar{\mathbf{x}}_d^T(\hat{\boldsymbol{\beta}}_{\hat{\theta}_d} - \hat{\boldsymbol{\beta}}_{0.5})$ at $q = \hat{\theta}_d$. It is convenient to re-scale the elements $\hat{\mathbf{u}}^{MQ}$ so that they have mean equal to zero.
- (Step 2) Construct the vector $\hat{\mathbf{u}}^{MQ*} = \{\hat{u}_1^{MQ*}, \dots, \hat{u}_D^{MQ*}\}^T$, whose elements are the components of the set $srswr(\{\hat{u}_1^{MQ}, \dots, \hat{u}_D^{MQ}\}, D)$.
- (Step 3) Generate a bootstrap population U^* of N independent bootstrap Bernoulli realisations made up of D areas with area d of size N_d, and with bootstrap Bernoulli realisation y_{dj}^* in area d taking the value 1 with probability

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

- (Step 4) Calculate the bootstrap population parameters $\bar{y}_d^*$, $d = 1, \dots, D$.
- (Step 5) Extract a sample s^* of size n from the bootstrap population U^* using the same sample design as that used to obtain the original sample and calculate the bootstrap M-quantile predictor $\hat{\bar{y}}_d^{MQ*}$, $d = 1, \dots, D$.
- (Step 6) Repeat steps 2-5 B times. In the bth bootstrap replication, let $\bar{y}_d^{*(b)}$ be the quantity of interest for area d and let $\hat{\bar{y}}_d^{MQ*(b)}$ be its corresponding M-quantile estimate.
- (Step 7) The NPB estimator of the MSE of $\hat{\bar{y}}_d^{MQ}$ is

$$mse^{NPB}(\hat{\bar{y}}_d^{MQ}) = B^{-1}\sum_{b=1}^{B}\left(\hat{\bar{y}}_d^{MQ*(b)} - \bar{y}_d^{*(b)}\right)^2. \tag{A1}$$

Random effect block bootstrap (REBB) procedure

The steps in the REBB bootstrap are as follows.

- (Step 1) Calculate D vectors of marginal residuals $\mathbf{r}_d^{MQ} = (r_{dj}^{MQ}) = \mathbf{x}_{dj}^T(\hat{\boldsymbol{\beta}}_{\hat{\theta}_d} - \hat{\boldsymbol{\beta}}_{0.5}),\ j = 1, \dots, n_d,\ d = 1, \dots, D$, re-scaling the elements of the vector $\mathbf{r}_d^{MQ}$ so that they have mean equal to zero.

- (Step 2) Construct the individual bootstrap errors for the N_d population units in area d as $\mathbf{r}_d^{MQ*} = (r_{dj}^{MQ*}) = srswr(\mathbf{r}_{h(d)}^{MQ}, N_d)$ where $h(d) = srswr(\{1, \ldots, D\}, 1)$.
- (Step 3) Generate a bootstrap population U^* of N independent bootstrap Bernoulli realisations made up of D areas with area d of size N_d, and with bootstrap Bernoulli realisation y_{dj}^* in area d taking the value 1 with probability

$$p_{dj}^* = \frac{\exp\{\mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}}_{0.5} + r_{dj}^{MQ*}\}}{1 + \exp\{\mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}}_{0.5} + r_{dj}^{MQ*}\}}, \; j = 1, \ldots, N_d.$$

- (Step 4) Calculate the bootstrap population parameters $\bar{y}_d^*$, $d = 1, \ldots, D$.
- (Step 5) Extract a sample s^* of size n from the bootstrap population U^* using the same sample design as that used to obtain the original sample and calculate the bootstrap M-quantile predictor $\hat{\bar{y}}_d^{MQ*}, \; d = 1, \ldots, D$.
- (Step 6) Repeat steps 2-5 B times. In the bth bootstrap replication, let $\bar{y}_d^{*(b)}$ be the quantity of interest for area d and let $\hat{\bar{y}}_d^{MQ*(b)}$ be its corresponding M-quantile estimate.
- (Step 7) The REBB estimator of the MSE of $\hat{\bar{y}}_d^{MQ}$ is

$$mse^{REBB}(\hat{\bar{y}}_d^{MQ}) = B^{-1} \sum_{b=1}^{B} \left(\hat{\bar{y}}_d^{MQ*(b)} - \bar{y}_d^{*(b)} \right)^2. \tag{A2}$$

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.

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. DOI:10.1080/10618600.2012.681237.

CHAMBERS, R., CHANDRA, H., SALVATI, N. & TZAVIDIS, N. (2009). Outlier robust small area estimation. In *Working Paper Series*. Wollongong: Centre for Statistical and Survey Methodology, University of Wollongong, pp. 1–40.

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

COPAS, J. B. (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society. Series B* **50**, 225–265.

FARRELL, P., MACGIBBON, B. & TOMBERLIN, T. (1997). Bootstrap adjustments for empirical bayes interval estimates of small area proportions. *Canadian Journal of Statistics* **25**, 75–89.

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.

HOROWITZ, J. (1992). A smoothed maximum score estimator for binary response model. *Econometrica* **60**, 505–531.

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.

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

KORDAS, G. (2006). Smoothed binary regression quantiles. *Journal of Applied Econometrics* **21**, 387–407.

MACGIBBON, B. & TOMBERLIN, T. (1989). Small area estimates of proportions via empirical bayes techniques. *Survey Methodology* **15**, 33–56.

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.

MANSKI, C. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* **3**, 205–228.

MANSKI, C. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics* **32**, 65–108.
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., SAEI, A. & LOMBARDÍA, M. (2007). Small area estimates of labour force participation under a multinomial logit mixed model. *Journal of the Royal Statistical Society, Series A* **70**, 265–283.
NANDRAM, B., SEDRANSK, J. & PICKLE, L. (1999). Bayesian analysis of mortality rates for u.s. health service areas. *Sankhya, B* **61**, 145–165.
NEWEY, W. & POWELL, J. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55**, 819–847.
NOH, M. & LEE, Y. (2007). Robust modeling for inference from generalized linear model classes. *Journal of the American Statistical Association* **102**, 10259–1072.
PRASAD, N. & RAO, J. (1990). The estimation of mean squared error of small-area estimators. *Journal of the American Statistical Association* **85**, 163–171.
PREGIBON, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics* **38**, 485–498.
PREISSER, J. & QAQISH, B. (1999). Robust regression for clustered data with application to binary responses. *Biometrics* **55**, 574–579.
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: 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.
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.
TZAVIDIS, N., MARCHETTI, S. & CHAMBERS, R. (2010). Robust estimation of small-area means and quantiles. *Australian and New Zealand Journal of Statistics* **52**, 167–186.