Semiparametric small area estimation for binary outcomes with applications to unemployment and poverty estimation

Ray Chambers
Centre for Statistical and Survey Methodology,
University of Wollongong, New South Wales 2522, Australia

Nicola Salvati
Dipartimento di Economia e Management
Università di Pisa, Via Ridolfi, 10 - I56124 Pisa, Italy

Nikos Tzavidis
Southampton Statistical Sciences Research Institute,
University of Southampton, Southampton SO17 1BJ, UK

Summary. A new semiparametric and robust approach for model-based small area prediction for discrete outcomes is proposed. This approach is using extensions of M-quantile regression for estimating small area proportions where a popular alternative approach is to use a predictor based on a generalised linear mixed model. The paper motivates the proposed methodology by means of two applications. The first is concerned with the estimation of unemployment levels for Local Authorities in the UK. The second investigates the estimation of the proportion of poor households in Local Labour Systems in the Tuscany region of Italy. In both applications the accuracy of direct estimates is poor and the proposed model-based methodology produces estimates with clearly improved coefficients of variation, which also compete with alternative model-based estimators. Two estimators of the prediction mean squared error are proposed, one using Taylor linerization and another using block bootstrap.

Keywords: EU-SILC; M-estimation; Generalized linear mixed model; Quantile regression; Robust inference; UK LFS

1. Introduction

Decision-makers tasked with devising and implementing policies to maximum effect need as much information as possible and at disaggregated geographical levels. 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 and 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 and Rao, 1990). An alternative approach to small area prediction is to use M-quantile regression (Breckling and Chambers, 1988) to characterise area heterogeneity (Chambers and Tzavidis, 2006). Unlike traditional prediction with mixed models, the M-quantile approach is semiparametric and automatically allows for outlier robust prediction. However, many survey

E-mail: ray@uow.edu.au
E-mail: salvati@ec.unipi.it
E-mail: n.tzavidis@soton.ac.uk

variables are categorical in nature and are therefore not suited to standard SAE methods based on LMMs. Two important discrete survey variables are unemployment and poverty, defined for the purposes of this paper as income deprivation.

The demand for small area estimates of labour force activity in the UK is described in a letter by the Librarian of the House of Commons to the Head of the Office for National Statistics (ONS) Labour Market Division. The letter identifies the need and stresses the importance for producing a common set of key labour force indicators at disaggregated geographical levels that can be used by Members of the House of Commons. Responding to this need, the Office for National Statistics in the UK has recognised that sample sizes at the level of Unitary Authorities/Local Authority Districts (UALADs) may not be sufficient to meet the 20% Coefficient of Variation (CV) threshold (ONS, 2004) necessary for publication of direct estimates of labour force activity. For this reason, ONS established a project aiming at investigating the use of model-based SAE methodology for producing reliable estimates of unemployment levels and rates for UALADs. This work resulted in the development of methodology that now allows ONS to publish estimates with improved accuracy, which are now are accredited as National Statistics.

Regarding poverty estimation, the Lisbon European Council (March 2000), assisted by the Nice European Council (December 2000) and by the Gothenburg Council (2001), agreed to put in place an EU strategy by the year 2010 aiming at making a decisive impact on the battle against poverty in European Union Countries. A first towards achieving this goal requires estimates of poverty levels at the small area level. However, only very recently National Statistical Institutes in Europe have started investigating methodologies for producing small area estimates of income poverty. In Italy, a target geography for producing poverty estimates is defined by Local Labour Systems (LLSs) but the sample sizes at this level do not allow direct estimation with acceptable CVs.

One option for small area prediction in the case of discrete outcomes is to adopt a Hierarchical Bayes approach (Malec et al., 1997; Nandram et al., 1999) or to use Empirical Bayes (MacGibbon and Tomberlin, 1989; Farrell et al., 1997). Alternatively, if a frequentist approach is preferred, one can follow Jiang and 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). For example, the ONS methodology for estimating unemployment levels and rates is based on a binary logistic GLMM with UALADs random effects. Large deviations from the expected response as well as outlying points in the space of the explanatory variables (leverage points) are known to have a large influence on classical maximum likelihood inference based on generalised linear and generalised linear mixed models (GLMs and GLMMs). This has lead to the development of robust methods for fitting these models (Pregibon, 1982; Preisser and Qaqish, 1999; Cantoni and Ronchetti, 2001; Noh and Lee, 2007). Sinha (2004) proposes a robust Monte Carlo Newton-Raphson method of estimation, which can be considered as a modification of the Monte Carlo Newton-Raphson method of McCulloch (1997). To obtain a Monte Carlo version of the robust maximum likelihood estimate, similarly to McCulloch (1997), the author presents a Metropolis algorithm to approximate the posterior distribution of the random effects. Maiti (2001) considered the development of robust methods for GLMM-based small area prediction using a Hierarchical Bayes approach to fitting a GLMM based on an outlier-robust normal mixture prior for the random effects.

To the best of our knowledge, there is no existing methodology for robust small area prediction for discrete outcomes in the frequentist framework. In this paper we therefore propose a new approach to SAE for discrete data based on M-quantile modelling. This is achieved by extending 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 and 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 achieved in the presence of both misclassification and measurement error.

Starting with the target of producing small area estimates of unemployment levels in UALADs in the UK and income poverty in Italy, two data sources are used for motivating the potential benefit of using robust prediction. These are the Labour Force Survey (LFS) in the UK and the European Survey of Income and Living Conditions (EU-SILC) in Italy. For the case of estimating unemployment levels for UALADs in the UK, our research thesis is that the proposed robust small area methodology offers reliable and in some cases improved estimates to the ones that are currently published. For the case of estimating poverty levels for LLSs in the Italy, our research thesis is that the proposed methodology offers estimates that are more accurate than direct estimates and competitive to alternative model-based estimates.

The structure of the paper is as follows. In Section 2 we review the current industry standard for estimating a small area proportion using a GLMM. The data we will use for small area estimation in this paper come from the UK LFS and the EU-SILC from Italy. In Section 3 we present the specification of a small area working model and exploratory analysis of these two datasets that aims at motivating the use of the proposed methodology. Section 4 describes the methodological contribution of this paper based on the extension of the M-quantile regression approach to binary data. In Section 5 we explore the use of the M-quantile approach for small area prediction with discrete outcomes. In this Section we further propose analytic and bootstrap estimators of the corresponding mean squared error (MSE). In Section 6 we use the M-quantile approach to estimate: (1) levels of ILO unemployment in 406 UALADs of the UK; and (2) the proportion of poor households (below the poverty line) in the 57 Local Labour Systems (LLSs) of the Tuscany region in Italy. In Section 7 we present results from model-based simulation studies aimed at assessing the robustness of the different small area predictors considered in this paper under a range of misspecification scenarios. Finally, in Section 8 we summarise the main findings.

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 (in this paper a discrete random 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 and 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

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

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

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

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 $\mathbf{\Sigma}_u$. The matrix $\mathbf{\Sigma}_u$ 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 Conditional Expectation Predictor (CEP) for the area d proportion $\bar{y}_d$ under (2),

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

where $\hat{\mu}_{dj} = \exp\{\hat{\eta}_{dj}\}(1+\exp\{\hat{\eta}_{dj}\})^{-1}$, $\hat{\eta}_{dj} = \mathbf{x}_{dj}^T\hat{\boldsymbol{\beta}} + \mathbf{z}_d^T\hat{\mathbf{u}}_d$, $\hat{\boldsymbol{\beta}}$ 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 $\mathbf{z}_d$ is a vector $(0, 0, \ldots, 1, \ldots, 0)$ with value 1 in the d-th position, in which case the $\mathbf{u}_d$ are scalar small area effects. We refer to (3) in this case as a 'random intercepts' CEP. For more details on this predictor, including estimation of its MSE, see Saei and Chambers (2003), Jiang and Lahiri (2006) and González-Manteiga et al. (2007). Note, however, that (3) is not the proper Empirical Best Predictor by Jiang (2003) because the author develop the Prasad-Rao type frequentist's alternative to the already existing hierarchical Bayes methods under the GLMMs. Under this approach the Empirical Best Predictor for $P(y_{dj} = 1|\mathbf{u}_d)$ is given by

$$\exp\{\mathbf{x}_{dj}^T\hat{\boldsymbol{\beta}}\}\frac{E[\exp\{(y_{d\cdot} + 1)\mathbf{\Sigma}_u\zeta - (n_d + 1)\log(1 + \exp\{\hat{\eta}_{dj}\})\}]}{E[\exp\{y_{d\cdot}\mathbf{\Sigma}_u\zeta - n_d\log(1 + \exp\{\hat{\eta}_{dj}\})\}]} \tag{4}$$

where $y_{d\cdot} = \sum_{j \in s_d} y_{dj}$ and the expectations are taken with respect to $\zeta \sim N(\mathbf{0}, \mathbf{I})$. The proper EBP does not have closed form and needs to be computed by numerical approximations. For this reason, the CEP version (3) is used in practice as is the case with the small area estimates of Labour Force activity currently produced by ONS in the UK. 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. Moreover, estimation can be numerically demanding. 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 and Chambers (2003). In the empirical results reported in Section 7, we used function `glmer` for parameter estimation.

3. Data sources, model specification and diagnostics

In this Section we describe the data available for performing small area estimation and also present diagnostics for the GLMM model. These diagnostics allow us to motivate the use of the alternative semiparametric methodology we propose in this paper.

3.1. The UK LFS

The data are coming from the UK LFS carried out by ONS in 2000. The LFS is a survey of households living at private addresses in the UK. Its purpose is to provide information on the UK labour market which can then be used to develop, manage, evaluate and report on labour market policies. From 1992 the sample in UK was increased to cover about $50,000$ households every quarter enabling quarterly publication of LFS estimates. In one of the applications of this paper we are interested in estimating unemployment levels for 406 UALADs in the UK. We use the ILO unemployment definition and data from a sample of about $169,000$ people aged 16 and over. ONS considers an estimate to be publishable if its coefficient of variation is less than 20%. According to this rule direct estimates can be published only for 75 out of the 406 UALADs with the data from 2000. For this reason the use of small area estimation appears to be appropriate in this case. The application of small area methodologies requires at the first stage the estimation of the model using the survey data. At the second stage the estimated model parameters are combined with population information for predicting the level of unemployment in each UALAD. The covariates we use in our working models are specified by prior studies of small area estimation of labour force characteristics in the UK (ONS, 2006) and confirmed by the analysis of deviance from fitting a binary GLMM with Normal random effects to the sample data. These covariates are: sex-age category of an individual (6 categories corresponding to female/male 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, Bailey et al., 2000) and total of registered unemployed in the sex-age group for the UALAD. The estimated model parameters and the resulting test statistics are presented in Table 1. The estimated model parameters are in the expected direction. For example, controlling for the effects of other explanatory variables and unobserved heterogeneity, the odds of being unemployed for young females are higher than the odds of any other age-sex group. The ratio unemployed over inactive and employed for sex-age=6 (men over 40) is about 0.09 times sex-age=1 (women aged between 16 and 25). When we move from the group of women over 40 (sex-age=3) to men aged between 16 and 25 (sex-age=4), there is a reduction of the probability to be unemployed. We can conclude that the probability to be unemployed decreases as age increases and this probability for men is less than that for women. The estimated σ_u and the value of the likelihood ratio test are shown in the last row of Table 1. The variance of the random effects is significant suggesting UALAD level heterogeneity in the levels of unemployment.

Figure 1 shows the normal probability plot of estimated UALAD random effects obtained by fitting a logistic mixed model to the sample data. This plot indicates some departures from normality in particular at the tails of the distribution. Furthermore, the distribution of level 1 Pearson residuals is skewed indicating the presence of potential influential observations in the data, with a number of large residuals ($|r_{dj}| > 2$). This is also confirmed by the application of a robust fitting method (Cantoni and Ronchetti, 2001) to this data set. In particular, the use of a robust GLM suggests that although most observations receive a weight of 1, about 3.5% receive weights less than 0.7. Therefore, using in prediction a model that bounds the influence of potentially outlying observations seems worthy of investigation.

3.2. The EU-SILC in Italy

In this second application we focus on estimating the number of poor households in each of the 57 LLSs of Tuscany, using the 2005 EU-SILC dataset. LLSs are defined as a collection of contiguous municipalities. These are areas in which most of the daily activity of the people who live and work in them takes place. Their definition is similar to that of the travel-to-work-areas widely used in US and UK territorial analyses. The data was collected by the National Statistical Institute in Italy

Table 1. Model fitting results UK Labour Force Survey data: '.' Significant at level 0.05, '$*$' significant at level 0.01, '$**$' significant at level 0.001, '$***$'.

Variable	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept	-3.60404	0.18124	-19.89	0.000000 $***$
registered unemployed	0.20762	0.02967	7.00	0.000000$***$
sex-age=2	-0.17508	0.05417	-3.23	0.001229$**$
sex-age=3	-1.06979	0.05003	-21.38	0.000000$***$
sex-age=4	-1.12496	0.04407	-25.53	0.000000$***$
sex-age=5	-1.62521	0.05081	-31.99	0.000000 $***$
sex-age=6	-2.33940	0.07604	-30.77	0.000000 $***$
government=2	-0.06045	0.08692	-0.70	0.486763
government=3	0.05140	0.11485	0.45	0.654493
government=4	-0.09204	0.13790	-0.67	0.504487
government=5	0.26696	0.09752	2.74	0.006191$**$
government=6	-0.05835	0.08521	-0.68	0.493448
government=7	0.08996	0.08818	1.02	0.307664
government=8	-0.11991	0.08387	-1.43	0.152812
government=9	-0.02315	0.09120	-0.25	0.799652
government=10	0.02623	0.09845	0.27	0.789926
government=11	0.09053	0.08705	1.04	0.298359
government=12	-0.14834	0.09464	-1.57	0.116990
socio-economic cluster=2	0.03377	0.06922	0.49	0.625705
socio-economic cluster=3	0.27070	0.07935	3.41	0.000647 $***$
socio-economic cluster=4	-0.04883	0.07690	-0.63	0.525454
socio-economic cluster=5	0.28525	0.07655	3.73	0.000194$***$
socio-economic cluster=6	0.05330	0.11659	0.46	0.647551
socio-economic cluster=7	0.33256	0.14546	2.29	0.022235 $**$
Parameter	Estimate	LR	$Pr(> chi2)$	
σ_u	0.18519	42.32	0.000000	

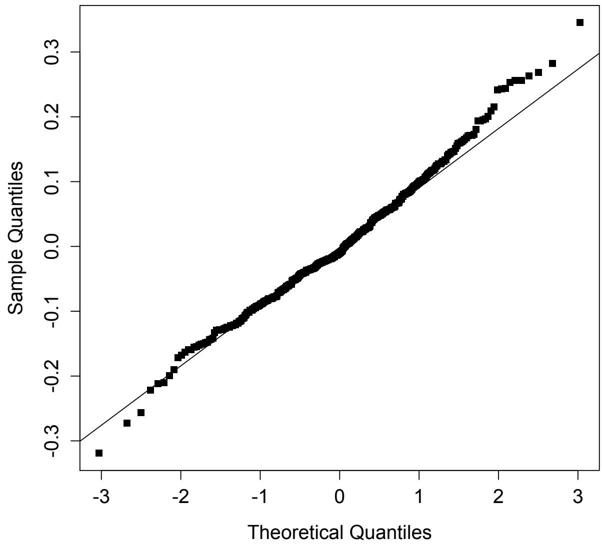

Fig. 1. Normal probability plot of estimated UALAD random effects for proportion of unemployed population aged 16 and over, based on a logistic mixed model fit to UK LFS data.

(ISTAT) as part of the 2005 EU-SILC survey. We used a sample of $1,560$ households across 29LLSs (out of 57) in Tuscany. The 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. 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 ($= 1$) if its equivalised income falls below a minimum level (the poverty line) necessary to meet basic food and non-food needs. The poverty line is defined as 60% of median equivalised income. In this paper the household equivalised income is computed by using the modified OECD scale (Hagenaars et al., 1994): it is calculated for each household as the household total disposable net income divided by the equivalised household size, which gives a weight of 1.0 to the first adult, 0.5 to other persons aged 14 or more and 0.3 to each child aged less than 14.

Direct estimates of the number of poor households within each LLS level have high variances, particularly for LLSs with small sample sizes. Moreover, direct estimates cannot be computed for areas with no sample. The covariates used in the working small area model for the binary outcome of being below the poverty line are ownership status (owner$= 1$), age of the head of the household, gender of the head of the household (female$= 0$) and interactions between these covariates. Table 2 presents the estimated model parameter from fitting a GLMM for the binary response. The estimates are in the expected direction with the odds of owners, for example, to be below the poverty line being lower than the odds of non-owners. We note that ownership status and gender are significant. Age appears to be non-significant, however, since the interaction term between ownership status$\times$age is

Table 2. Model fitting results EU-SILC data: '.' Significant at level 0.05, '$*$' significant at level 0.01, '$**$' significant at level 0.001, '$***$'.

Variable	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept	-0.08060	0.70509	0.11	0.909
ownership status	-2.73428	0.77693	-3.52	0.000$***$
age	-0.01090	0.01129	-0.97	0.334
gender	-1.43413	0.63847	-2.25	0.025 $*$
ownership status$\times$age	0.03472	0.01207	2.87	0.004 $**$
ownership status$\times$gender	0.58208	0.43060	1.35	0.176
gender$\times$age	0.00262	0.00966	0.27	0.786
Parameter	Estimate	LR	$Pr(> chi2)$	
σ_u	0.47949	14.67	0.000110	

significant, we decided to leave the age effect in the model. The results show that owner old man has less probability to be poor. The last row of the table gives the estimate of σ_u and the value of the likelihood ratio test (LRT) statistic suggesting that the variance of the random effects is significant. Population data for these covariates were taken 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 and this has been shown by work conducted as part of the SAMPLE project (Small Area Methods for Poverty and Living Conditions Estimates, URL: http://www.sample- project.eu/). See Fabrizi et al. (2013) for a detailed discussion.

Figure 2 shows the distribution of the LLS-level residuals and the normal probability plot of the of the LLS-level residuals obtained from the logistic mixed model fit to the EU-SILC data. The departures from normality are even more pronounced in the case of the EU-SILC data. 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 by fitting the model using a robust method (Cantoni and Ronchetti, 2001) and note that although most observations receive a weight of 1 in this fit, there are about 8.5% of the overall sample that receive a weight of less than 0.7. This is confirmed by the Cooks distances, reported in Figure 2, which indicate that there are outliers in the data. Note that in this Figure values greater than $4/n \approx 0.003$ are considered as potential outliers. These diagnostics provide evidence for the presence of outlying observations in this dataset.

4. M-quantile regression for binary outcomes

In this Section we present the extension of linear M-quantile regression to binary data. Since M-quantile regression does not depend on how areas are specified, in the notation we use in this Section we drop subscript d.

4.1. M-quantile regression for a continuous response

M-quantile regression (Breckling and 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

$$\int \psi_q\Big(\frac{Y - Q_q}{\sigma_q}\Big) dF(Y) = 0, \tag{5}$$

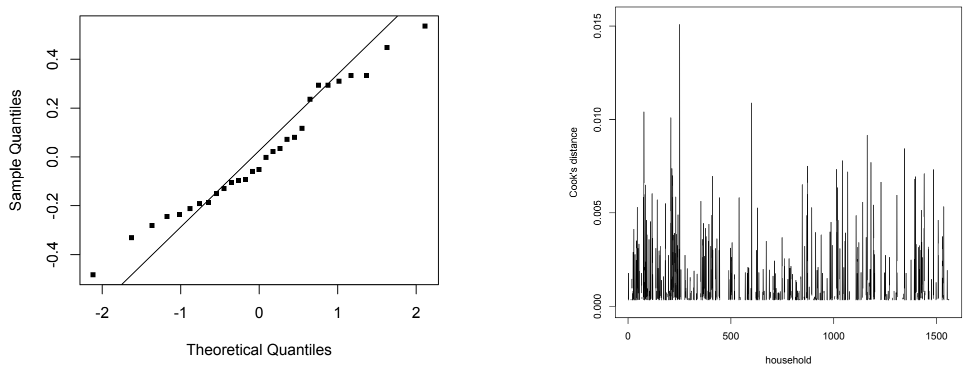

Fig. 2. Model fit diagnostics for a logistic mixed model fit to the EU-SILC data: normal probability plot of the LLS-level residuals (left-hand-side plots) and Cooks distances (right-hand-side plots).

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 and Powell, 1987), and when $\psi(t) = sgn(t)$ we obtain the standard quantile of order q (Koenker and Bassett, 1978).

Breckling and 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

$$Q_q(\mathbf{X}; \psi) = \mathbf{X}\boldsymbol{\beta}_q. \tag{6}$$

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

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

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 (7) using standard iteratively re-weighted least squares (IRLS).

4.2. M-quantile regression for binary outcomes: 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 (5) when Y is binary, with $P(Y = 1) = p$.

In this case (5) 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 (6) 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, we propose to replace (6) by the linear logistic specification

$$Q_q(\mathbf{x}_j;\psi) = \frac{\exp(\mathbf{x}_j^T\boldsymbol{\beta}_q)}{1+\exp(\mathbf{x}_j^T\boldsymbol{\beta}_q)}. \tag{8}$$

For estimating $\boldsymbol{\beta}_q$ we consider the extension to the M-quantile case of the Cantoni and 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:

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

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 (9). 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 and Ronchetti (2001) refer to the solution to (9) as the Huber quasi-likelihood estimator. When ψ is the identity function, (9) reduces to the usual maximum likelihood estimating equations for a GLM.

In the case of binary outcomes, the estimating equation (9) can be extended for obtaining the M-quantile fit by applying the idea of asymmetric weighting of Pearson residuals that was also used in the linear case. In particular, the estimating equations can be re-written as

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

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

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

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

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. The estimating equation (10) can be solved numerically using a Fisher scoring procedure to obtain an estimate $\hat{\boldsymbol{\beta}}_q$ of $\boldsymbol{\beta}_q$.

Note that when $q = 0.5$, (10) reduces to (9). Moreover, (7) is a special case of (10) 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 count random variable (Tzavidis et al., 2013).

Assuming that ψ is a continuous monotone non-decreasing function, a first order approximation to the variance of (10) is given by

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

Detailed expressions of these quantities are given in Appendix I. *R* routines for estimation and inference using M-quantile regression with binary and count data are available from the authors.

4.3. *Links with the econometric literature*

The estimating equation approach described in the previous Section does not apply to standard quantile regression for binary data. Quantile regression for binary data has been developed in the econometric literature using a latent variable concept. However, as we now show, our proposed approach and the econometric approach are very closely related, since the econometric approach can be shown to be equivalent to the solution of an estimating equation analogous to (7). Since we confine ourselves to standard quantiles in this Section, 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

$$\hat{\boldsymbol{\beta}}_q = \max_{\|\mathbf{b}=1\|} n^{-1}\sum_{j=1}^{n}\{y_j - (1-q)\}I(\mathbf{x}_j^T\mathbf{b} > 0) \tag{12}$$

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 (12) reduces to

$$\hat{\boldsymbol{\beta}}_q = \min_{\|\mathbf{b}=1\|} n^{-1}\sum_{j=1}^{n}\Big[qI\{y_j \geqslant I_j(\mathbf{b})\} + (1-q)I\{y_j < I_j(\mathbf{b})\}\Big]|y_j - I_j(\mathbf{b})|. \tag{13}$$

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 (7) 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 (12) has a solution.

A smoothed version of (12) has been proposed by Horowitz (1992) as having better finite sample properties:

$$\hat{\boldsymbol{\beta}}_q = \max_{\|\mathbf{b}=1\|} n^{-1} \sum_{j=1}^{n} \{y_j - (1-q)\} F(\sigma_n^{-1} \mathbf{x}_j^T \mathbf{b}), \tag{14}$$

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 (13) allow us to write (14) 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 (8). 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 (7), with $Q_q(\mathbf{x}_j;\psi)$ defined by (8) and subject to the restriction $\|\boldsymbol{\beta}_q = 1\|$.

4.4. Links with statistical literature

Efron (1992) proposed an alternative approach to modelling the conditional distribution of a count outcome given the covariates using asymmetric maximum likelihood (AML) estimation. As Machado and 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. Efrons' approach results in estimates of the conditional location that is similar to conditional expectiles proposed by Newey and Powell (1987). Efrons' method can be extended to model the conditional distribution of a binary outcome given the covariates. Using the binomial deviance, the AML estimate $\hat{\boldsymbol{\beta}}_w$ for $\boldsymbol{\beta}$ can defined as

$$\hat{\boldsymbol{\beta}}_w = \arg\max_{\mathbf{b}} n^{-1} \sum_{j=1}^{n} [y_j \log(\mu_j(\mathbf{b})) + (1-y_j)\log(1-\mu_j(\mathbf{b}))] w^{I\{y_j > \mu_j(\mathbf{b})\}}, \tag{15}$$

where $\mu_j(\mathbf{b}) = \exp\{\mathbf{x}_j^T\mathbf{b}\}\Big(1+\exp\{\mathbf{x}_j^T\mathbf{b}\}\Big)^{-1}$. From (15), by vector differentiation, the following estimating equation is obtained:

$$n^{-1} \sum_{j=1}^{n} \Big[(y_j - \mu_j(\mathbf{b}))\mathbf{x}_j^T\Big] w^{I\{y_j > \mu_j(\mathbf{b})\}} = \mathbf{0}. \tag{16}$$

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 are those obtained by Efrons' asymmetric maximum likelihood estimation for a specific choice of w. In particular, setting the tuning constant equal to a large value, (10) can be written as:

$$\Psi(\boldsymbol{\beta}_q) := n^{-1} \sum_{j=1}^{n} \Big\{ (y_j - Q_q(\mathbf{x}_j;\psi)) w_{jq} \mathbf{x}_j^T \Big\} = \mathbf{0}, \tag{17}$$

where $w_{jq} = \Big[qI\{y_j > Q_q(\mathbf{x}_j;\psi)\} + (1-q)I\{y_j \leqslant Q_q(\mathbf{x}_j;\psi)\}\Big]$. This weight can be also written as $w_{jq} = \Big[\Big(\frac{q}{1-q}\Big)I\{y_j > Q_q(\mathbf{x}_j;\psi)\} + I\{y_j \leqslant Q_q(\mathbf{x}_j;\psi)\}\Big]$. Setting $w = \frac{q}{(1-q)}$ in Efrons' estimating equation (15) results in estimates that are equivalent to those obtained from our proposed estimating equation (17).

5. Robust prediction of small area proportions using M-quantile regression

Many survey variables are binary and there is a growing demand for reliable small area estimates based on such variables. From now on therefore we focus on using M-quantile regression models for binary outcomes with the aim of obtaining small area estimates of proportions.

5.1. The M-quantile small area population model

Random effects models use random area effects to account for between-area variation. The M-quantile approach to small area prediction suggests a completely different approach to capturing area heterogeneity. To start with, the population model is specified at the unit level. Define q_{dj} such that $y_{dj} = Q_{q_{dj}}(\mathbf{x}_{dj};\psi)$.Under the logit transformation, the population model is defined by

$$Q_{q_{dj}}(\mathbf{x}_{dj};\psi) = \exp\{\mathbf{x}_{dj}^T\boldsymbol{\beta}_{q_{dj}}\}\Big(1+\exp\{\mathbf{x}_{dj}^T\boldsymbol{\beta}_{q_{dj}}\}\Big)^{-1}.$$

Chambers and Tzavidis (2006) used the term M-quantile coefficients for q_{dj}. The variability in q_{dj} reflects variability at the unit level. If clustering exists, then units in the same clusters (clusters) will have similar M-quantile coefficients and different from those of units that belong to other clusters (areas). A cluster specific M-quantile coefficient is then defined as $\theta_d = E[q_{dj}|d]$.

5.2. Point estimation

Provided there are sample observations in area d, 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 and Tzavidis (2006), the M-quantile predictor of the average $\bar{y}_d$ in small area d is

$$\hat{\bar{y}}_d^{MQ} = N_d^{-1}\Big\{\sum_{j\in s_d} y_{dj} + \sum_{j\in r_d}\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_{dj};\psi)\Big\}. \quad (18)$$

When Y is binary, and we model its regression M-quantile of order q via (8), 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 (18). 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 extensions of the M-quantile coefficient concept to binary Y before we consider inference based on (18).

5.3. 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

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

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 7, 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 3. 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 3 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.

5.4. Mean squared error estimation

In this section we propose a MSE estimator for (18) based on the linearisation approach set out in Chambers et al. (2013). 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 (18) is then

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

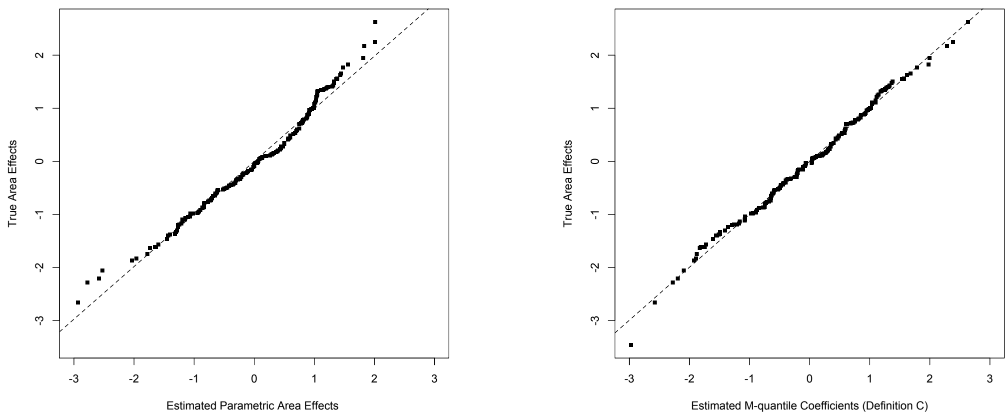

Fig. 3. Estimated area effects vs. true area effects (left plot) and estimated M-quantile coefficients vs. true area effects (right plot) in a Monte-Carlo simulation with $D = 200$ and $n_d = 25$.

which can be estimated by

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

Here $\widehat{Var}(\hat{\boldsymbol{\beta}}_{\hat{\theta}_d})$ is a sandwich-type estimator that can be estimated using the expressions in Appendix I 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):

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

with corresponding plug-in estimator

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

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

$$mse^A(\hat{\bar{y}}_d^{MQ}) = \widehat{Var}(\hat{\bar{y}}_d^{MQ}) + \{\widehat{Bias}(\hat{\bar{y}}_d^{MQ})\}^2. \quad (19)$$

In the development above we make the standard assumption that a consistent estimator of the MSE of a linear approximation to the small area estimator of interest can be used as its MSE estimator. As noted by Harville and Jeske (1992), such an approach will not generally be consistent, and the resulting MSE estimator can be biased low. The MSE estimator ignores the contribution to the mean squared error from the estimation of the area level M-quantile coefficients by $\hat{\theta}_d$. This is a linearisation assumption since for large overall sample sizes the contribution to the overall mean squared error of (18) arising from the variability of $\hat{\theta}_d$ is of smaller order of magnitude than the prediction variance of (18). As a consequence, (19) will tend to be almost unbiased. However, the potential underestimation of the MSE of (18) implicit in (19) needs to be balanced against the bias robustness of this MSE estimator under misspecification of the second order moments of y, and may well lead to (19) being preferable to other MSE estimators based on higher order approximations that depend on the model assumptions being true (Chambers et al., 2013).

We also propose a bootstrap-based method for estimating the MSE of (18), which is a form of a random effects block bootstrap. In order to save space, the computational details the bootstrap procedure are set out in the Appendix II. Here we summarise the main characteristics of the method. The block bootstrap, see Chambers and Chandra (2013), 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 BB) for estimating the distribution of the M-quantile predictor (18) 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.

6. Applications

6.1. Estimating levels of ILO unemployment for UALADs in the UK

In this Section we use the model specification of Section 3.1 and the CEP and M-quantile predictors for estimating the level of ILO unemployment for UALADs in the UK. For assessing the resulting estimates we note that model-based small area estimates should be consistent with corresponding unbiased direct estimates but more precise. In addition, given the model diagnostics presented in Section 3.1, we may expect that the proposed robust approach may offer some gains in efficiency over the CEP predictor. Figure 4 maps the estimated levels of ILO unemployment for UALADs in the UK in 2000 using three small area predictors. The emerging patterns of unemployment for UALADs in the UK produced by using the proposed methodology are consistent with the patterns reported by ONS (2006). Hubs of higher unemployment in 2000 are located in UALADs in London, parts of Wales, North-east and North-west England and Scotland. However, as the preceding analysis showed, the proposed methodology may offer more accurate estimates.

We now present further evidence from internal validation of our estimates. Figure 5 presents M-quantile estimates of the total number of unemployed against corresponding direct estimates for each UALAD. We note that M-quantile estimates appear to be generally consistent with the direct estimates with the correlation between the two sets of estimates being 0.78.

In order to assess the potential gains in precision from using model-based estimates instead of direct estimates, we examine the distribution of the ratios of the estimated CVs of the direct

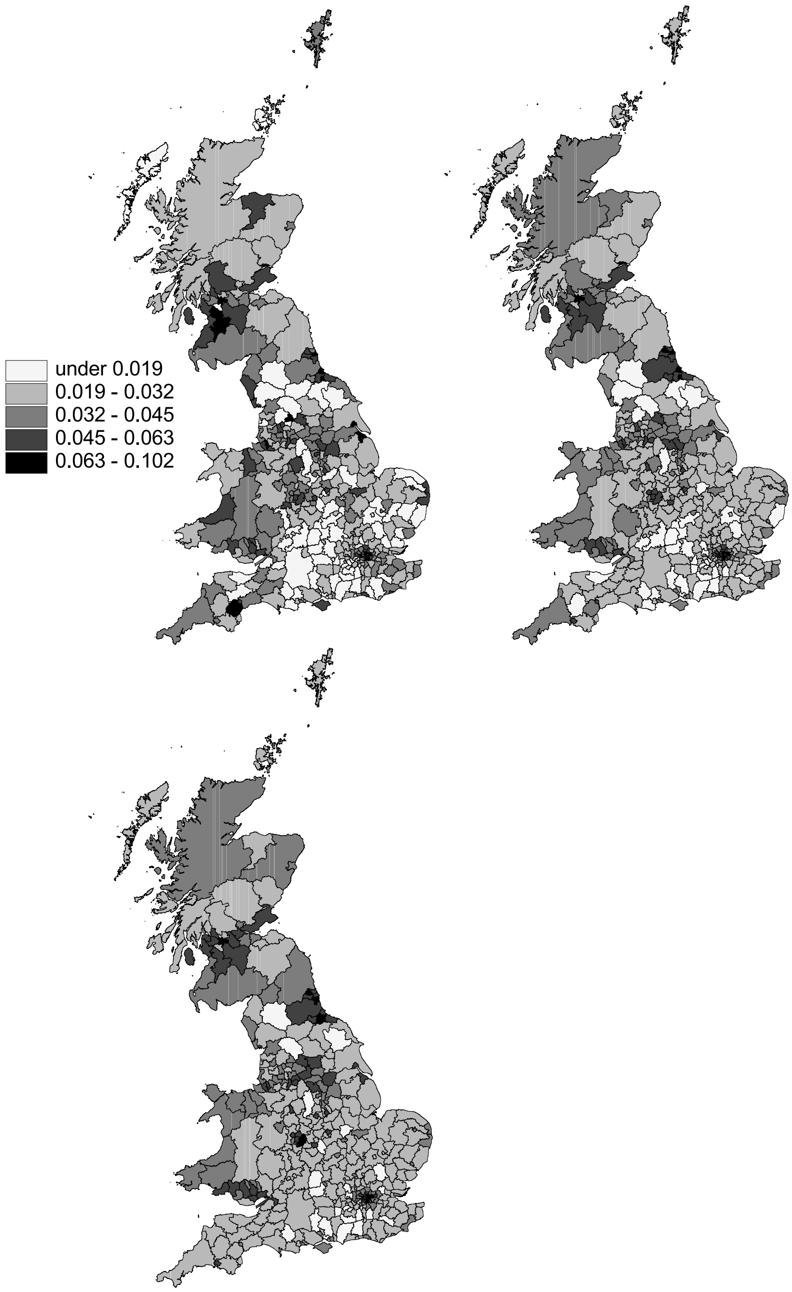

Fig. 4. Maps of the estimates of levels of ILO unemployment for UALADs in the UK in 2000: Direct (top left), CEP (top right) and MQ (bottom left) estimates.

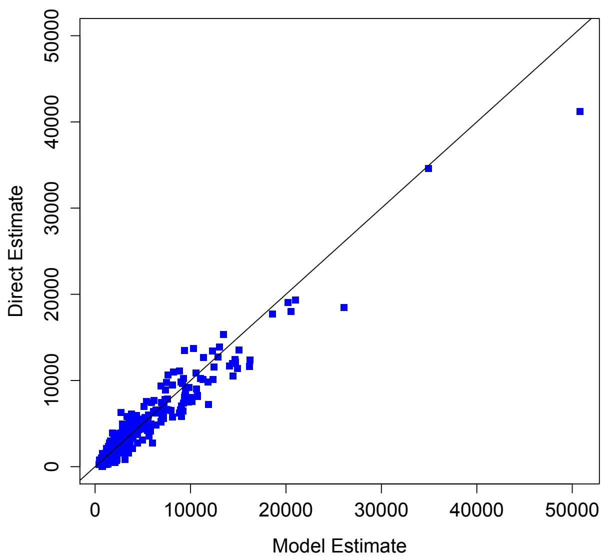

Fig. 5. Numbers of unemployed people aged 16 and over in UALADs in the UK in 2000: Model-based M-quantile estimates versus corresponding direct estimates.

and the model-based estimates. A value greater than 1 for this ratio indicates that the estimated CV of the model-based estimate is smaller than that of the direct estimate. Figure 6 shows the relationship between these ratios and the number of unemployed people in each UALAD. Two sets of ratios are plotted - those corresponding to CEP estimates (red) and those corresponding to the M-quantile estimates (blue). Note that the CV for the M-quantile estimates is calculated as $[mse^{BB}(\hat{\bar{y}}_d^{MQ})]^{1/2}/\hat{\bar{y}}_d^{MQ}$, while that for the CEP estimates is calculated as $[mse^{EBP}(\hat{\bar{y}}_d^{CEP})]^{1/2}/\hat{\bar{y}}_d^{CEP}$, where $mse^{CEP}(\hat{\bar{y}}_d^{CEP})$ is obtained using the bootstrap procedure proposed in González-Manteiga et al. (2007). Figure 6 shows that the estimated CVs of the M-quantile and CEP 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 CEP estimates, indicating better accuracy. 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.

6.2. Estimating poverty levels in LLSs of Tuscany

In this Section we derive estimates of the proportion of households below the poverty line in LLSs of Tuscany in 2005 using the model specifications of Section 3.2. Figure 7 maps the small area poverty estimates. These maps indicate that the LLSs in the north-western part and northern part of the region of Tuscany, corresponding to the LLSs in the province of Massa-Carrara and in the northern part of the provinces of Lucca and Prato, are characterised by the highest estimates of poor households. Hence, these areas can be considered as the most critical in the region. On the other hand, the LLSs with the lowest proportions of poor households are found in the provinces of

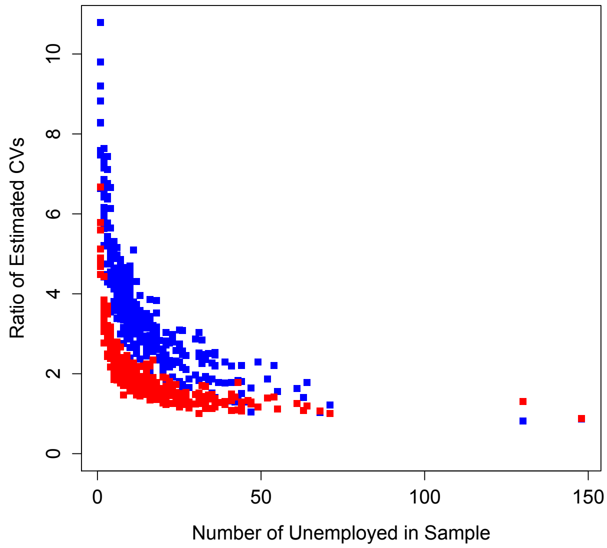

Fig. 6. Ratio of estimated coefficients of variation of direct estimates to M-quantile (blue) and CEP (red) estimates of total number of unemployed for each UALAD.

Florence, Siena and Arezzo, in the central-eastern part of Tuscany. One thing to note is that direct estimates are not produced for all LLSs because of the absence of sample units from these LLSs in the EU-SILC. For those cases model-based estimation is the only option.

Validating the small area estimates is a challenging task due to the lack of official estimates at this level of geography. Some internal diagnostics are presented below. The correlation between the direct and M-quantile estimates for the sampled LLSs is 0.97, while the corresponding correlation for the CEP estimates is 0.95. As Figure 8 shows, M-quantile estimates are consistent with direct estimates. The left plot in Figure 9 presents ratios of the estimated CVs for the direct and model-based estimates of numbers of poor households and the right plot shows the estimated CVs of these estimates across LLSs (in percentage terms). In both plots blue indicates M-quantile estimates and red indicates CEP estimates. The solid black line in the right plot refers to the direct estimates. The efficiency gains by using model-based estimates over direct estimates are not as striking as in the previous application but are still substantial, particularly for LLSs with a small number of sampled households. Generally, the M-quantile estimates have a smaller estimated CV than corresponding CEP estimates, with the most striking differences found for the non-sampled LLSs (located to the right of the vertical line).

7. Model-based simulations

The validity of model-based inference depends on the validity of the model. The preceding analysis relies on the analysis of one sample, which makes it difficult to generalise our findings. In this Section we empirically evaluate the properties of small area predictors and corresponding MSE

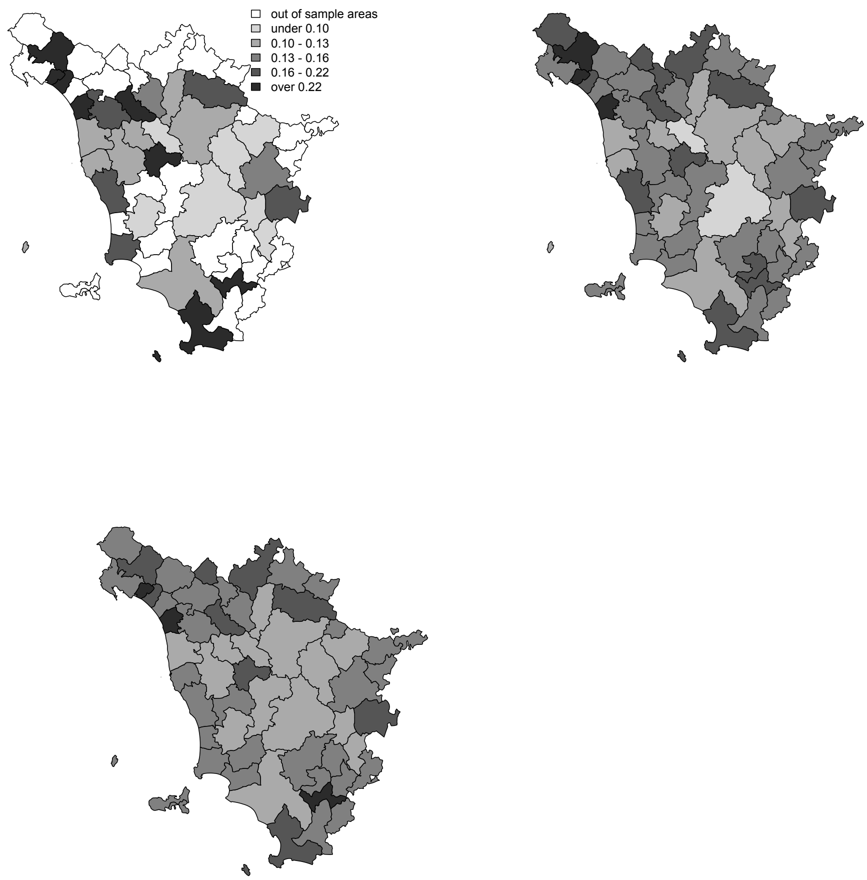

Fig. 7. Maps of the estimated proportion of poor households for LLSs in Tuscany in 2005: Direct (top left), CEP (top right) and MQ (bottom left) estimates.

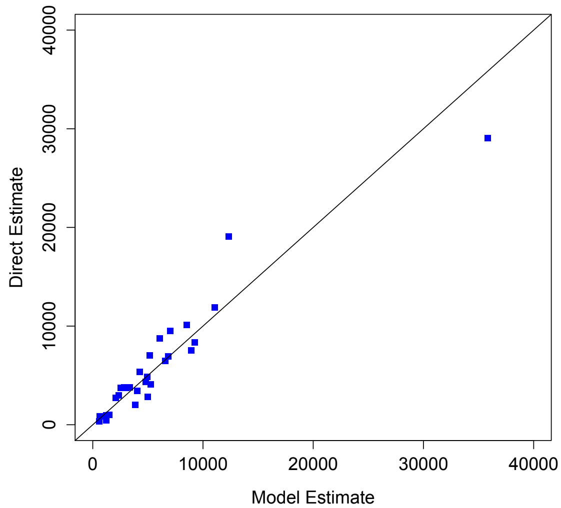

Fig. 8. Model-based M-quantile estimates of the number of households in poverty in sampled LLSs compared with corresponding direct estimates.

estimators. This is done by means of a sensitivity analysis to departures from the GLMM model assumptions using Monte Carlo simulation. In particular, population data are first generated under different scenarios for departures from the assumptions of the GLMM model with samples then being selected from this simulated population. Estimates of small area proportions and the corresponding MSE are computed, using the samples, and empirically assessed.

Two different M-quantile versions of (18) 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 CEP (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 the bootstrap method described in Section 5.4. 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 CEP were estimated using the function `glmer` in *R*.

In each simulation we generated $N = 5,000$ population values of X and Y in $D = 50$ small areas with $N_d = 100$, $d = 1, \ldots, D$. Individual x_{dj} values were drawn independently at each simulation as $Uniform(a_d, b_d)$, for $a_d = -1$ and $b_d = d/4$, $d = 1, \ldots, D$, $j = 1, \ldots, N_d$. 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

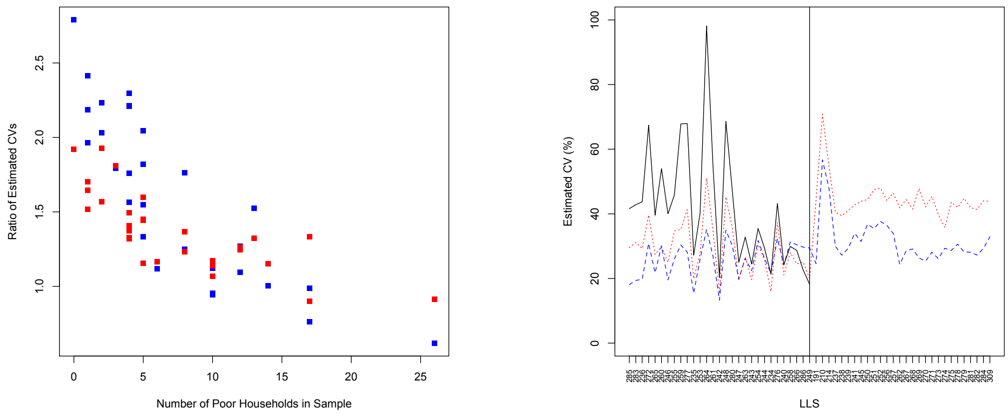

Fig. 9. The left plot shows the ratio of estimated CVs of direct estimates to M-quantile (blue) and CEP (red) estimates of the proportion of poor households for each LLS. The right plot shows the estimated CVs for the direct (solid line) and the model-based estimates. Estimated CVs for the M-quantile predictor shown by the dashed blue line and estimated CVs for the CEP shown by the dashed red line.

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 CEP and the direct estimator were used to estimate the small area proportions $\bar{y}_d,\ d = 1, \ldots, D$.

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 3, where we see that claims in the literature (Chambers and Tzavidis, 2006) about the superior outlier robustness of the M-quantile predictor compared with the CEP and the Expectile predictor certainly hold true in these simulations. In particular, under the (0) scenario the CEP 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 no notable difference between CEP, M-quantile and Expectile under the (0) scenario, while under the (M) scenario the M-quantile predictor appears to be superior.

In order to evaluate the performances of the MSE estimators for M-quantile predictor ($c = 1.345$) proposed in Section 5.4 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

Table 3. Model-based simulation results: Predictors of small area proportions.

Predictor/Scenario	$n_d = 10$ (0)	$n_d = 10$ (M)	$n_d = 20$ (0)	$n_d = 20$ (M)
	Median values of Bias			
CEP	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			
CEP	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

Table 4. 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)
$mse^A(\hat{\bar{y}}_d^{MQ})$	-5.36	-6.05	24.34	25.30	95	94
$mse^{BB}(\hat{\bar{y}}_d^{MQ})$	-0.73	-0.91	13.66	12.90	95	95
			$D = 100$			
$mse^A(\hat{\bar{y}}_d^{MQ})$	-5.81	-6.91	25.12	24.51	95	94
$mse^{BB}(\hat{\bar{y}}_d^{MQ})$	-0.85	-0.86	12.08	11.49	95	95

calculated as well as its linearisation MSE estimator (19) and the bootstrap MSE estimator BB using 100 bootstrap iterations. The performance of these MSE estimators for each scenario is presented in Table 4 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 (19) these intervals were defined by the small area estimate plus or minus twice the value of the square root of (19). For BB these intervals were based on the 2.5 and the 97.5 percentiles of the corresponding bootstrap distribution. Examination of the results in Table 4 shows that both MSE estimation methods tend to be biased low, but all generate nominal 95 per cent confidence intervals with acceptable coverage. Overall, the BB estimator seems preferable because it shows smaller bias and better stability than the linearisation-based estimator (19).

8. Final remarks

Small area prediction for discrete outcomes is an important and challenging problem. In this paper we propose a new approach for small area estimation of binary outcomes. The proposed methodology uses an extension of M-quantile regression for discrete data, which is then adapted for small area estimation. By construction, the proposed approach is outlier robust. The benefits of using the new method are illustrated in two applications. In both applications the results show that the proposed methodology produces model-based estimates that are consistent but more efficient than direct estimates and competitive to alternative model-based estimates. We further present two approaches for estimating the MSE of the M-quantile model-based predictor, one based on a linearisation argument and the other based on block bootstrap. The MSE estimators provide acceptable coverage performance in our simulations, with the block bootstrap being perhaps preferable because of its stability and simplicity.

An obvious extension of the development set out in this paper is the development of M-quantile versions of GLMs for count data. Although we briefly describe an extension to GLM-type modelling for a count outcome, we do not explore the behaviour of corresponding small area estimators. This is an area of current research. We also do not explore the extension of M-quantile modelling to multi-category outcomes, which remains an open problem.

Appendix I

A first order approximation to the variance of (10) is given by (11) where

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

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

$a_j^2(\boldsymbol{\beta}_q)$ 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

$$\widehat{Var}(\hat{\boldsymbol{\beta}}_q) = n^{-1}\hat{\mathbf{B}}^{-1}(\hat{\boldsymbol{\beta}}_q)\widehat{Var}\{\Psi(\hat{\boldsymbol{\beta}}_q)\}[\hat{\mathbf{B}}^{-1}(\hat{\boldsymbol{\beta}}_q)]^T. \quad (20)$$

Appendix II

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.

Block bootstrap procedure

The steps in the BB bootstrap are as follows.

- (Step 1) Calculate D vectors of marginal residuals $\mathbf{r}_d^{MQ} = (r_{dj}^{MQ}) = \mathbf{x}_d^T(\hat{\boldsymbol{\beta}}_{q_{dj}} - \hat{\boldsymbol{\beta}}_{0.5})$, $j = 1, \ldots, n_d,\ d = 1, \ldots, D$, 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 BB estimator of the MSE of $\hat{\bar{y}}_d^{MQ}$ is

$$mse^{BB}(\hat{\bar{y}}_d^{MQ}) = B^{-1} \sum_{b=1}^{B} \left(\hat{\bar{y}}_d^{MQ*(b)} - \bar{y}_d^{*(b)} \right)^2. \tag{21}$$

References

Bailey, S., J. Charlton, G. Dollamore, and J. Fitzpatrick (2000). Families, groups and clusters of local and health authorities: revised for authories in 1999. *Population Trends 99*, 37–52.

Battese, G., R. Harter, and W. Fuller (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. and R. Chambers (1988). M-quantiles. *Biometrika 75*, 761–771.

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

Chambers, R. and H. Chandra (2013). A random effect block bootstrap for clustered data. *Journal of Computational and Graphical Statistics 22*, 452–470.

Chambers, R., H. Chandra, N. Salvati, and N. Tzavidis (2013). Outlier robust small area estimation. *Journal of the Royal Statistical Society. Series B 75*, Forthcoming.

Chambers, R. and N. Tzavidis (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.

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

Fabrizi, E., C. Giusti, N. Salvati, and Tzavidis (2013). Mapping average equivalized income using robust small area methods. *Papers in Regional Science Forthcoming*.

Farrell, P., B. MacGibbon, and T. Tomberlin (1997). Bootstrap adjustments for empirical bayes interval estimates of small area proportions. *Canadian Journal of Statistics 25*, 75–89.

González-Manteiga, W., M. Lombardía, I. Molina, D. Morales, and L. Santamaría (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.

Hagenaars, A., K. de Vos, and M. A. Zaidi (1994). *Poverty statistics in the late 1980s: Research* *based on micro-data*. Office for Official Publications of the European Communities, Luxembourg.

Harville, D. A. and D. R. Jeske (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association 87*, 724–731.

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

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

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

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

MacGibbon, B. and T. Tomberlin (1989). Small area estimates of proportions via empirical bayes techniques. *Survey Methodology 15*, 33–56.

Machado, J. and S. Silva (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., J. Sedransk, C. Moriarity, and F. LeClere (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.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized lin- ear mixed models. *Journal of the American Statistical Association 92*, 162–170.

Nandram, B., J. Sedransk, and L. Pickle (1999). Bayesian analysis of mortality rates for u.s. health service areas. *Sankhya, B 61*, 145–165.

Newey, W. and J. Powell (1987). Asymmetric least squares estimation and testing. *Econometrica 55*, 819–847.

Noh, M. and Y. Lee (2007). Robust modeling for inference from generalized linear model classes. *Journal of the American Statistical Association 102*, 1059–1072.

ONS (2006). *Model-based estimates of ILO unemployment for LAD/UAs in Great Britain. Guide for users*. Downloadable at http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/subnational-labour/model-based-estimates-of-ilo-unemployment-for-lad-uas-in-great-britain—guide-for-users.pdf.

Prasad, N. and J. Rao (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. and B. Qaqish (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. and R. Chambers (2003). Small area estimation under linear and generalized linear mixed models with time and area effects. In *S3RI Methodology Working Papers*, pp. 1–35. Southampton:

Southampton Statistical Sciences Research Institute.

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

Tzavidis, N., M. Ranalli, N. Salvati, E. Dreassi, and R. Chambers (2013). Poisson m-quantile regression for small area estimation. *Paper submitted and available from the authors upon request*.