Semiparametric small area estimation for binary outcomes with application to unemployment estimation for Local Authorities in the UK

Ray Chambers* Nicola Salvati† Nikos Tzavidis‡

Abstract

A new semiparametric and robust approach to small area estimation for discrete outcomes is proposed. The methodology represents an efficient and easily computed alternative to prediction using a generalised linear mixed model and is based on an extension of M-quantile regression. In addition, two estimators of the prediction mean squared error are described, one based on Taylor linearization and another based on block bootstrap. The proposed methodology is applied to UK annual Labour Force Survey data for estimating the proportion of the unemployed in Local Authorities in the UK. The properties of estimators are further empirically assessed in model-based simulations.

Keywords: M-estimation; M-quantiles; Generalized linear mixed model; Robust inference; UK Labour Force Survey.

1 Introduction

Decision-makers tasked with devising and implementing policies to maximum effect need information at disaggregated geographical levels. Such information can be obtained by producing small area estimates derived from data collected in national surveys. The UK is one of few countries in Europe where the National Statistics agency (the Office for National Statistics or ONS) regularly produces small area estimates which have also gained accreditation as National Statistics. An example of such accredited small area National Statistics is the annual set of unemployment estimates for Unitary Authorities/Local Authority Districts (hereafter referred to as UALADs) in the UK, which are derived from data collected in the UK Labour Force Survey, or UKLFS. The demand for small area estimates of labour force activity in the UK was recently highlighted in a letter written by the Librarian of the House of Commons to the

*National Institute for Applied Statistics Research Australia, University of Wollongong, New South Wales 2522, Australia, ray@uow.edu.au

†Dipartimento di Economia e Management, Università di Pisa, Via Ridolfi, 10 - I56124 Pisa, Italy, salvati@ec.unipi.it

‡Department of Social Statistics and Demography and Southampton Statistical Sciences Research Institute, University of Southampton, Southampton SO17 1BJ, UK, n.tzavidis@soton.ac.uk

Head of the ONS' Labour Market Division, where the importance of a common set of key labour force indicators at disaggregated geographical levels that can be used by Members of the House of Commons is emphasised. Responding to this need, the ONS has recognised that sample sizes for the UKLFS within UALADs are not sufficient to meet the 20% Coefficient of Variation (CV) threshold necessary for publication of direct survey estimates of labour force activity (ONS, 2006), and has implemented the use of model-based small area estimation (SAE) methodology for producing official annual unemployment estimates for UALADs.

The increasing demand for reliable small area statistics has led to the development of a number of efficient mixed model-based 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 continuously distributed variable (Battese et al., 1988; Prasad and Rao, 1990). An alternative approach to small area estimation is to use M-quantile regression models (Breckling and Chambers, 1988) to characterise between area variation (Chambers and Tzavidis, 2006). Unlike prediction based on mixed models, the M-quantile approach is semiparametric and automatically allows for outlier robust prediction. However, many survey variables, such as the International Labour Organisation (ILO) definition of unemployment, are categorical in nature and are therefore not suited to standard SAE methods based on LMMs.

The unemployment status of a person is a binary outcome. One option for small area prediction in the case of binary 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) or Jiang (2003) who propose an empirical best predictor (EBP) for a binary response. Ugarte et al. (2009) analyse the performance of several design-based, model-assisted, and model-based estimators for unemployment at the small area level by using different sources of auxiliary information. Molina et al. (2007) use a multinomial logit mixed model for labour force activity with random effects that are assumed to be the same across the categories of the response variable. López-Vizcaíno et al. (2013, 2014) propose a multinomial logit mixed model for small area estimation of a categorical response which they use to estimate labour force activity in Galicia, Spain. Their model allows for category-specific time and domain (area) random effects, which seems appropriate. The current ONS methodology for estimating unemployment levels and rates is based on a binary logistic mixed model with random effects specified at the level of UALADS.

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 Generalized Linear Models (GLMs) and Generalized Linear Mixed Models (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).

Tzavidis et al. (2014) propose a new semiparametric M-quantile approach to small area prediction for counts that extends the ideas of Cantoni and Ronchetti (2001) and Chambers and Tzavidis (2006). This predictor can be viewed as an outlier robust alternative to the more commonly used Conditional Expectation Predictor for counts that is based on a Poisson GLMM with Gaussian random effects. In this paper we propose a new robust approach to SAE for binary outcomes that is based on M-quantile modelling. In particular, we extend the M-quantile approach to SAE for continuous data (Chambers and Tzavidis, 2006) and for counts (Tzavidis et al., 2014; Chambers et al., 014b) to the case where the response is binary. Modelling the M-quantiles of a binary outcome presents more challenges than modelling the M-quantiles of a count outcome. A detailed account of these challenges is provided in the present paper. With the proposed approach random effects are avoided and between area variation in the response is characterised by variation in area-specific values of M-quantile indices. Furthermore, outlier-robust inference is achieved in the presence of both misclassification and measurement error.

To motivate the potential benefits from using the proposed methodology, we use data from the UKLFS that have the same structure as those used by Molina et al. (2007). However, the methodology presented in this paper is more aligned with the current ONS methodology, in the sense that we focus on estimating unemployment for UALADs in the UK, with the aim of demonstrating whether it is possible to produce reliable and in some cases improved estimates compared to the GLMM-based estimates that are currently published by the ONS.

The structure of the paper is as follows. In Section 2 we review the current approach to using a GLMM to estimate a small area proportion. In Section 3 we develop working models for SAE of unemployment based on data from the UKLFS. To motivate the need for robust estimation methods, particular emphasis in this Section is given to model diagnostics. Section 4 then describes M-quantile regression for binary data. The links between the statistical and the econometric literature on modelling quantiles of discrete outcomes are also discussed in this Section. This methodology is extended in Section 5 to an M-quantile approach to SAE for a binary outcome. In this Section we further propose analytic and bootstrap estimators of the mean squared error (MSE) of small area estimators. In Section 6 we use these SAE methodologies to estimate levels of ILO unemployment in the 406 UALADs of the UK in the year 2000. 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 model misspecification scenarios. In addition, the consistency of estimators is empirically evaluated. Finally, in Section 8 we summarise the main findings of the paper.

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 for population unit j in area d, and let $\mathbf{x}_{dj}$ denote a $p \times 1$ vector of unit level covariates (including an intercept). In general 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. However, in the important special case where the components of $\mathbf{x}_{dj}$ are all categorical, we note that application of the methods described in this paper only require that the area level tabulations of these variables be available. 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 $\delta_d; d = 1, ..., D$ 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 δ_d is then $E_s[\delta_d]$.

In many cases $\delta_d = N_d^{-1} \sum_{j \in U_d} f(y_{dj})$ where f is a known function. The minimum mean squared error predictor of δ_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 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 $\boldsymbol{\Sigma}_u$. This matrix is allowed to depend on parameters, 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. This model is widely used in applied work, and is the one currently used by the ONS for calculating estimates of annual UALAD unemployment. 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). Despite their attractive properties as far as modelling binary (and, more generally, discrete-valued) response variables are concerned, application of GLMMs in small area estimation is not straightforward since estimation of model parameters 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 by using an iterative procedure that combines Maximum Penalized Quasi-Likelihood (MPQL) estimation of $\boldsymbol{\beta}$ and $\mathbf{u}_d$ with REML estimation of the variance components (Saei and Chambers, 2003). In the empirical results reported in Section 7, we used the function `glmer` for fitting the GLMM defined by (1) and (2).

An alternative to (3) is the Empirical Best Predictor (EBP) of $P(y_{dj} = 1|\mathbf{u}_d)$, see Jiang (2003). This is given by

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

where $y_{d\cdot} = \sum_{j\in s_d} y_{dj}$ and the expectations are taken with respect to $\zeta \sim N(\mathbf{0}, \mathbf{I})$. This predictor does not have a closed form and can only be computed via numerical approximation. Computing EBPs is generally not straightforward, however, which is why National Statistical agencies like the ONS favour computation of an approximation like the CEP. It is our understanding that an approximation like the CEP is also used in Molina et al. (2007) and López-Vizcaíno et al. (2013, 2014). Nevertheless, in this paper we further include EBP estimation, for the UKLFS application, by using the computational tools developed in Burgard (2013).

3 Data sources, model specification and diagnostics for the UKLFS data

In this Section we describe the UKLFS data set that we use for illustrating the proposed methodology. We also present diagnostics from fitting a GLMM to this data. These diagnostics allow us to subsequently motivate the use of an alternative semiparametric methodology based on M-quantile models.

3.1 Data structure

The data that we use is a subset of the annual data set created by the ONS using UKLFS data from the year 2000. The UKLFS is a quarterly 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. In this paper we focus on using these data to estimate unemployment levels for the 406 UALADs in the UK in 2000. We use the ILO definition of unemployment and our sample consists of about $169,000$ individuals aged 16 and over. The ONS considers an estimate to be publishable if its estimated coefficient of variation is less than 20 per cent. With this rule, direct survey estimates can only be published for 75 out of the 406 UALADs given the data from 2000. Application of SAE methods based on the CEP, see (3), significantly increases this number. Note that this requires that an appropriate GLMM first be fitted to the survey data. Estimated model parameters for fixed and random effects are then combined with known population information for each UALAD in order to predict its level of unemployment. The covariates we considered in our working GLMM are all either categorical or correspond to sex-age by area counts derived from demographic and administrative data sources and are based on prior studies of small area labour force characteristics in the UK (Molina et al., 2007; ONS, 2006). These covariates are: sex-age category of an individual, with six categories corresponding to female/male and three age groups ($16-25$, $26-40$ and >40), government office region of the UALAD (twelve categories), ONS socio-economic classification of the UALAD (Bailey et al., 2000) and total number of registered unemployed in the sex-age group for the UALAD. The ONS socio-economic classification of the UALAD consists of the following seven categories: rural areas, urban fringe, coast and services, prosperous England, mining, manufacturing and industry, education centres and outer London and inner London. The total number of registered unemployed disaggregated by age and sex for each UALAD is available from regularly updated administrative data sources, and represents the most important contextual covariate used in the model. Government office region and ONS socio-economic classification are area-specific categorical variables with values assigned by the ONS.

We emphasise that all explanatory variables used in this working model for UALAD unemployment are categorical variables for which sex-age by area counts are available. This corresponds to the situation considered by Molina et al. (2007), and so model-fitting does not require access to Census micro-data. This access would be required however if the model were to include one or more person level covariates with values that varied within an age-sex by area group.

3.2 Model fitting and model diagnostics

We start by fitting a binary logistic GLMM with normally distributed random effects to the UKLFS data. The estimated model parameters and the corresponding test statistics are set out

in Table 1 and 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. Similarly, the odds of being unemployed for sex-age=6 (men over 40) is about 0.09 times the odds for 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 in the odds of being unemployed. We conclude that the odds of being unemployed decrease as age increases and are lower for men than for women.

The significance of the variance component quantifying the between UALAD heterogeneity in unemployment can be tested by using a Likelihood Ratio Test (LRT). The value of this test statistic is 42.32, with a p-value $1.22e-09$, which provides evidence of significant unobserved UALAD heterogeneity in the observed levels of UALAD unemployment in 2000. Given that this is a test of whether this variance component is zero, i.e. a value on the boundary of the parameter space, this p-value has been determined by using a $50:50$ mixture of a χ^2_0 and a χ^2_1 distribution (Self and Liang, 1987). The LRT is perhaps the most commonly used approach for testing the significance of variance components. Results reported in Berkhof and Snijders (2001) show that the LRT has good properties especially when there is a large number of level 2 units as is the case with the UKLFS dataset that we use in this paper.

Figure 1 shows the normal probability plot of estimated UALAD random effects (left plot) and of Pearson residuals (right plot) obtained from fitting the logistic GLMM to the UKLFS data. These plots indicate some departures from normality particularly in the tails of the distribution.This is confirmed by a Shapiro-Wilk normality test, which rejects the null hypothesis that the random effects and the Pearson residuals follow a normal distribution (p-values $= 0.01371$ and 0.0000, respectively). Furthermore, the distribution of the Pearson residuals indicates the presence of potentially influential observations defined by large Pearson residuals ($|r_{dj}| > 2$). One diagnostic tool for checking for the presence of such influential observations is based on the work of Cantoni and Ronchetti (2001). Following these authors, we first carry out a robust logistic GLM fit to the data of interest. This fit is defined by weights that modify the impact of sample outliers. Our results suggest that although most sex-age by area cells receive a weight of 1 in the robust fit, about 17% receive weights less than 1 indicating that these cells correspond to potentially influential observations. Note that there is no difference between the robust and non-robust fits when all weights are equal to 1. Figure 2 shows the distribution of the weights, by gender and age across UALADs, with values less than 1 generated by the robust GLM fit to the UKLFS data. We can see that there is a substantial number of potentially influential data values in every sex-age group. We conclude that developing a prediction approach that bounds the influence of such outlying observations seems worthy of investigation for the UKLFS data.

Table 1: Model fitting results for UKLFS 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 $**$
Variance Component	Estimate	LR		$Pr(>\chi^2)$
σ_u	0.18519	42.32		1.22e-09

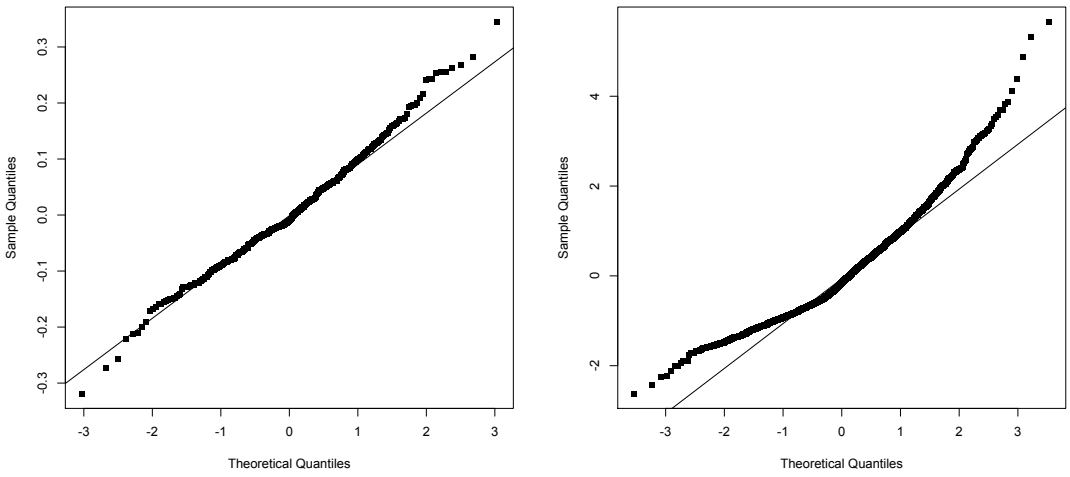

Figure 1: Normal probability plot of estimated UALAD random effects (left plot) and of level 1 Pearson residuals (right plot) based on a logistic GLMM fit to UKLFS data.

4 M-quantile regression for binary outcomes

The diagnostic work described in the preceding Section indicates that a prediction approach that bounds the influence of potentially outlying observations seems worthy of investigation when estimating UALAD unemployment from UKLFS data. In this Section we therefore describe an extension of the robust M-quantile regression modelling approach to binary data. Since M-quantile regression modelling does not depend on how areas are specified, we drop the subscript d in the notation used in this Section.

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

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

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

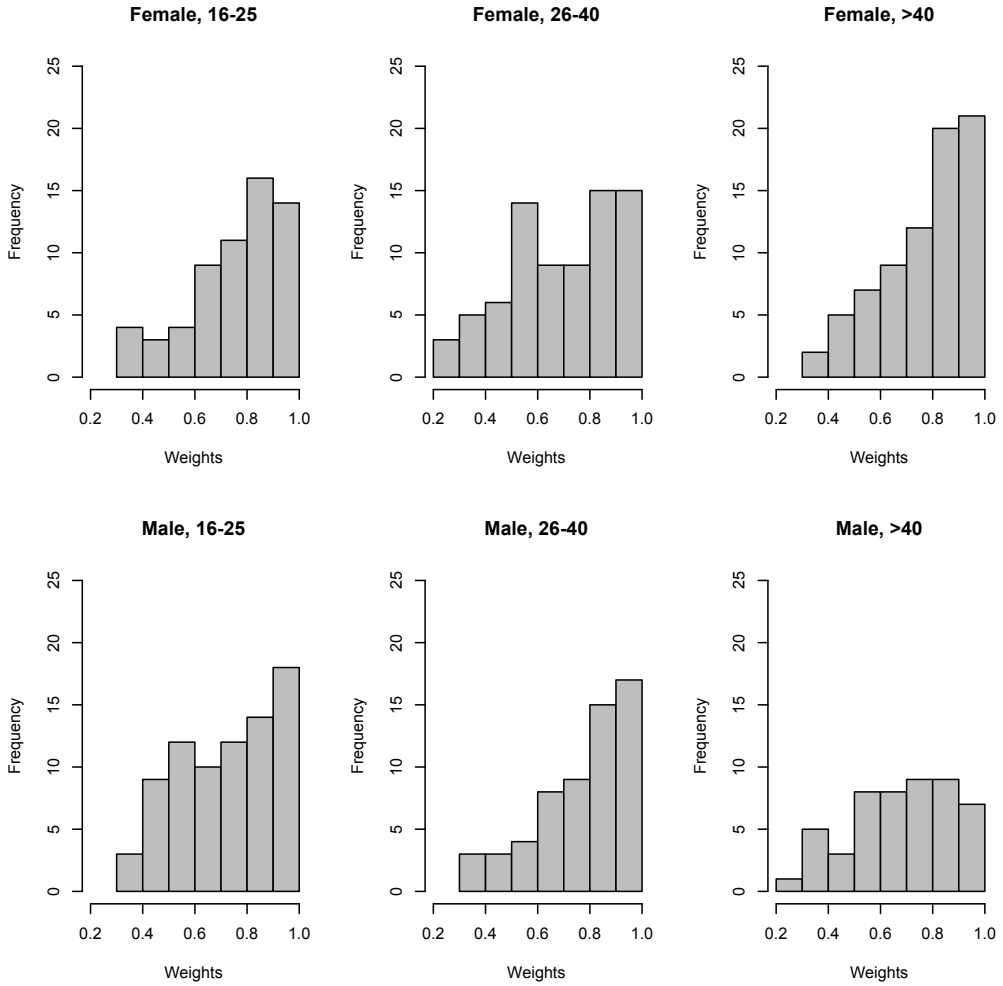

Figure 2: Histograms of the gender by age distributions of the weights from the robust logistic GLM fit to UKLFS data across UALADs. Weights equal to one have been excluded from the plot.

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 easily 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:

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

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 to obtain the M-quantile fit by applying the same asymmetric weighting of Pearson residuals as in the linear

case. In particular, the estimating equations (9) can be re-written as

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

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 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). Furthermore, this estimating equation approach applies quite generally. For example, it can be used when Y is a count (Tzavidis et al., 2014; Chambers et al., 014b).

At this point we must emphasise again that under the M-quantile modelling approach we follow the tradition of M-estimation (Huber, 1981) and do not make explicit distributional assumptions, e.g. assuming a Bernoulli distribution for Y. For fitting the M-quantile model we use a logit link function and a quasi-likelihood estimator (McCullagh and Nelder, 1989) that requires only the specification of working mean and the variance functions for Y. No explicit distributional assumption is made as is the case under maximum-likelihood estimation. The influence function ψ defining the M-quantile of interest is chosen in order to bound the impact of influential points in the space of the outcome and/or the explanatory variables. We then use the estimating function of Cantoni and Ronchetti (2001) to define the quasi-likelihood estimator for the parameters of the assumed mean and variance functions.

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

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

Detailed expressions for 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, which 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 latter 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 depend on the outcome of a continuously distributed latent variable. In particular, we assume that the observed value y_j is generated by an unobserved (latent) real value y_j^* in the sense that $y_j = 1$ when $y_j^* > 0$. Let $Q_q^*(\mathbf{x}_j)$ denote the conditional quantile function of this latent variable. Since $y_j = I(y_j^* > 0)$ is a monotone 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 $Q_q(\mathbf{x}_j) = I(\mathbf{x}_j^T\boldsymbol{\beta}_q > 0)$ 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 the statistical literature

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

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

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:

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

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 Efron's 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:

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

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 $w_{jq} = \left[\left(\frac{q}{1-q}\right)I\{y_j > Q_q(\mathbf{x}_j;\psi)\} + I\{y_j \leqslant Q_q(\mathbf{x}_j;\psi)\}\right]$. Setting $w = \frac{q}{(1-q)}$ in Efron's estimating equation (15) therefore results in estimates that are the same as 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 model

Mixed models use random area effects to account for between-area variation in SAE. The M-quantile approach to SAE is based on a completely different way of modelling between-area heterogeneity. To start, the population model is specified (and fitted) at the unit level and is free of any small area geography. Next, define q_{dj} such that $y_{dj} = Q_{q_{dj}}(\mathbf{x}_{dj};\psi)$. That is, q_{dj} is a random index that varies between zero and one. Assuming a logit specification, the population M-quantile model for q_{dj} (and hence y_{dj}) is then 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) refer to the q_{dj} as the M-quantile coefficients. Their variability reflects variability at the unit level. If clustering exists, population units in the same cluster (or small area) will have similar M-quantile coefficients and these will be different from those of units that belong to other clusters (or areas). An area d-specific M-quantile coefficient is then defined as $\theta_d = E[q_{dj}|d]$, where the expectation is conditional on the distribution of the random indices q_{dj} within area d.

5.2 Point estimation

Chambers and Tzavidis (2006) define the empirical value $\hat{q}_{dj}$ of the random index q_{dj} as the solution to $y_{dj} = \hat{Q}_{\hat{q}_{dj}}(\mathbf{x}_{dj};\psi)$ and refer to this value as the estimated M-quantile coefficient of y_{dj}. Provided there are sample observations in area d, and non-informative sampling method has been used to obtain them, an estimate $\hat{\theta}_d$ of the area d-specific M-quantile coefficient θ_d is the sample average of the estimated M-quantile coefficients for that area, otherwise we set $\hat{\theta}_d = 0.5$. The corresponding M-quantile predictor of the average $\bar{y}_d$ in small area d is

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

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 the estimated area level M-quantile coefficient $\hat{\theta}_d$, 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.

Small area estimation via GLMMs is based on predicted area effects. Similarly, small area estimation using the binary M-quantile model proposed in this paper is based on estimated area level M-quantile coefficients. A natural question to ask then concerns the strength of the relationship between the predicted area effects and the estimated area level M-quantile coefficients. Some empirical evidence for this relationship can be obtained by conducting a simulation experiment to show how the predicted area effects defined by the `glmer` function in `R` and the estimated area level M-quantile coefficients based on Definition C are related to true area effects. The simulated dataset was 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 $Normal(0,1)$.

The average correlation between the true area effects and the predicted area effects, over $1,000$ simulations, was 0.89, and the corresponding correlation between the true area effects and the estimated area level M-quantile coefficients was 0.80. These results suggest that estimated area level M-quantile coefficients are comparable to predicted area effects computed 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 an M-quantile approach to 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. (014a). 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

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

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 a sandwich-type estimator that can be calculated 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

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

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. As noted earlier, the MSE estimator (19) ignores the contribution to the mean squared error from 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). 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., 014a).

A bootstrap-based method for estimating the MSE of (18) can also be implemented. This is based on the random effects block (REB) bootstrap of Chambers and Chandra (2013). In order to save space, the computational details for this bootstrap procedure are set out in the Appendix II. Here we summarise the main characteristics of the method, which is a robust alternative to the parametric bootstrap for clustered data. The REB bootstrap is free of both the distribution and the independence assumptions of the parametric bootstrap and is consistent when the mixed model assumption is valid. In particular, it preserves area effects by bootstrap resampling within areas. Here we adapt this procedure in order to estimate the distribution of the M-quantile predictor (18). This is accomplished 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 in order 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 by using Bernoulli simulation.

6 Estimating levels of ILO unemployment for UALADs in the UK

In this Section we present the results from the application of the CEP and M-quantile approaches to estimating the level of ILO unemployment for UALADs in the UK. We also consider EBP estimation using the computational algorithms in Burgard (2013). In order to assess the resulting estimates we use a set of diagnostics based on the requirement 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, we expect that the proposed robust M-quantile approach should offer some efficiency gains over alternative predictors.

The GLMM is fitted using the `glmer` function in the package `lme4` of `R`. The M-quantile model is fitted in `R` using a modified version of the robust GLM fitting procedure described in Cantoni and Ronchetti (2001). Details of this algorithm are available from the authors of this paper on request. The influence function ψ used in the M-quantile model is the Huber Proposal 2 function with tuning constant $c = 1.345$. The specification of the working model is set out in Section 3 and was validated via a robust stepwise fitting procedure based on the Huber quasi-deviance (Cantoni and Ronchetti, 2001). The analysis of deviance is reported in Table 2 and shows that the model covariates are all highly significant predictors of unemployment.

Table 3 presents results from a robust fit of a Binomial GLM with a logit link function to the UKLFS data and the corresponding non-robust fit of the same model. These results indicate that there are some differences between the robust and non-robust model fits. This was expected given the diagnostic analysis presented in Section 3.

Table 2: Analysis of quasi-deviance table for the M-quantile model at $q = 0.5$. '.' Significant at level 0.05, '$*$' significant at level 0.01, '$**$' significant at level 0.001, '$***$'.

Covariates	pseudo df	df	χ^2 value	p-value
Null	168977			
registered unemployed	168976	1	894.8	$0.0000 * **$
registered unemployed, sex-age	168971	5	1793.8	$0.0000 * **$
registered unemployed, sex-age, government	168960	11	65.0	$0.0000 * **$
registered unemployed, sex-age, government, socio-economic cluster	168954	6	53.5	$0.0000 * **$

Figure 3 maps the estimated levels of ILO unemployment for UALADs in the UK in 2000 using three SAE methodologies namely, direct estimation, CEP estimation and M-quantile estimation. The emerging patterns of unemployment for UALADs in the UK produced by the proposed M-quantile methodology are consistent with the patterns reported by ONS (2006). Clusters of higher unemployment in 2000 are located in UALADs in parts of London, South Wales, North-east and North-west England and Scotland.

Although maps offer an effective tool for summarising small area estimates, they do not answer a fundamental question. How good are the estimates produced by using the proposed binary M-quantile model? In order to answer this question, we note that model-based estimates

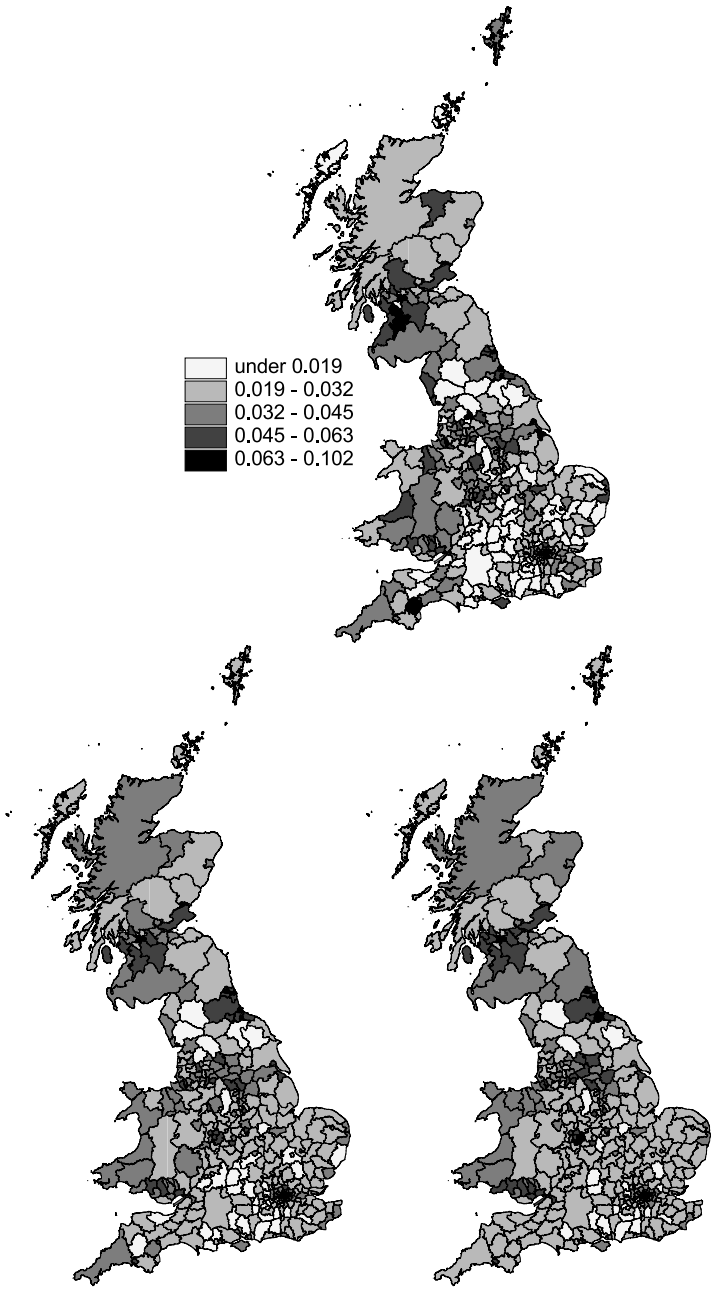

Figure 3: Maps of estimated levels of ILO unemployment for UALADs in the UK in 2000: Direct (top), CEP (bottom left) and M-quantile (bottom right) estimates.

Table 3: Estimated coefficients of the robust GLM logistic model fit ($\boldsymbol{\beta}_{Rob}$) and the non-robust GLM logistic model fit ($\boldsymbol{\beta}$) .

Variable	$\boldsymbol{\beta}_{Rob}$			$\boldsymbol{\beta}$		
	Estimate	Std. Error	$Pr(>\|z\|)$	Estimate	Std. Error	$Pr(>\|z\|)$
Intercept	-3.35478	0.14593	0.000000	-3.32730	0.14063	0.000000
registered unemployed	0.16948	0.02295	0.000000	0.16018	0.02214	0.000000
sex-age=2	-0.24560	0.05344	-0.000004	-0.20798	0.05155	0.000054
sex-age=3	-1.05495	0.04733	0.000000	-1.01950	0.04592	0.000000
sex-age=4	-1.16072	0.04459	-0.000000	-1.13803	0.04327	0.000000
sex-age=5	-1.63163	0.05179	0.000000	-1.63127	0.05034	0.000000
sex-age=6	-2.41673	0.07444	0.000000	-2.39604	0.07110	0.000000
government=2	-0.07659	0.07680	0.318627	-0.05160	0.07369	0.483805
government=3	0.05644	0.09093	0.534794	0.07652	0.08791	0.384015
government=4	-0.04934	0.09970	0.620659	-0.03599	0.09679	0.709987
government=5	0.25743	0.07798	0.000962	0.25886	0.07563	0.000620
government=6	-0.05733	0.07083	0.418323	-0.05510	0.06857	0.421684
government=7	0.07565	0.07063	0.284165	0.08899	0.06837	0.193101
government=8	-0.15685	0.07396	0.033941	-0.11721	0.07082	0.097904
government=9	-0.04077	0.07894	0.605478	-0.02246	0.07592	0.767321
government=10	0.02560	0.08229	0.755685	0.02512	0.07973	0.752755
government=11	0.03893	0.07127	0.584881	0.04193	0.06900	0.543398
government=12	-0.19680	0.07380	0.007660	-0.14136	0.07100	0.046483
socio-economic cluster=2	0.03396	0.06126	0.579389	0.04577	0.05872	0.435708
socio-economic cluster=3	0.24173	0.06856	0.000422	0.24935	0.06580	0.000151
socio-economic cluster=4	-0.07265	0.06958	0.296433	-0.07014	0.06648	0.291409
socio-economic cluster=5	0.31740	0.06489	0.000001	0.32286	0.06242	0.000000
socio-economic cluster=6	0.09617	0.09133	0.292351	0.08438	0.08827	0.339108
socio-economic cluster=7	0.40772	0.11271	0.000297	0.38995	0.10920	0.000356

should be (i) 'close' to the direct estimates and (ii) more precise than direct estimates. Following Brown et al. (2001), we assess (i) by computing a goodness of fit (GoF) diagnostic. This is based on the idea that if model-based estimates are 'close' to the small area value of interest, then unbiased direct estimates can be considered as random variables whose expected values are equal to the values of the corresponding model-based estimates. The GoF diagnostic is computed as the value of the following Wald statistic for each model based estimator:

$$W = \sum_d \Big\{ \frac{(\hat{\bar{y}}_d^{direct} - \hat{\bar{y}}_d^{model})^2}{[\widehat{Var}(\hat{\bar{y}}_d^{direct}) + \widehat{MSE}(\hat{\bar{y}}_d^{model})]} \Big\}.$$

The realised value of W can then be compared against the 0.95 quantile of a χ^2 distribution with $D = 406$ degrees of freedom, i.e. 453.98. Note that the $\widehat{MSE}$ of the M-quantile estimates is calculated using the REB bootstrap of Section 5.4 while the corresponding value for the CEP estimates is calculated by using the parametric bootstrap proposed in González-Manteiga et al. (2007). The values of the GoF are 412.64 for M-quantile estimates and 209.52 for CEP estimates. That is, both sets of model-based estimates are not statistically different from the

direct estimates. Figure 4 compares the M-quantile estimates of the total number of unemployed for each UALAD with the corresponding direct estimates. We note that the M-quantile estimates appear to be generally consistent with the direct estimates, with the correlation between the two sets of estimates being 0.78. The corresponding correlation between the direct estimates and the CEP estimates is 0.87. The higher correlation between the CEP and the direct estimates, compared to correlation between the M-quantile and the direct estimates, is consistent with the GoF values reported above.

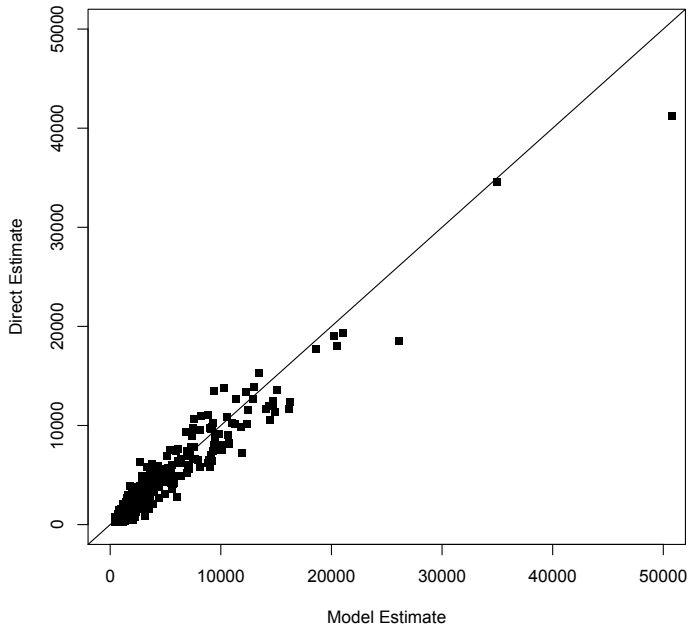

Figure 4: Numbers of unemployed people aged 16 and over in UALADs in the UK in 2000: M-quantile estimates versus corresponding direct estimates.

In order to assess (ii) above, i.e. the potential gains in precision from using model-based estimates (either CEP or M-quantile) instead of the direct estimates, we examine the distribution of the ratios of the estimated CVs of the direct and the model-based estimates for the UKLFS data. 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 5 shows the relationship between these ratios and the number of unemployed people in the UKLFS sample in each UALAD. Two sets of ratios are plotted - those corresponding to the CEP estimates (black) and those corresponding to the M-quantile estimates (gray). Figure 5 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 potentially better accuracy. One source for the differences between the CVs of the CEP and M-quantile estimates may be the presence of influential points identified in the diagnostic analysis reported in Section 3. From Figure 5 we also see that the differences between the CEP and the M-quantile estimates become more

evident as the number of unemployed in the sample decreases. In those cases there may be additional variability associated with the prediction of the random effects that contributes to the higher CVs of the CEP estimates.

One could also use an EBP-based approach to estimate UALAD unemployment. But, as noted earlier, computing EBPs is not a straightforward task, and the SAE methodology currently in use by the ONS is based on the CEP. Nevertheless, we calculated EBP estimates of UALAD unemployment in UKLFS 2000 by using the numerical algorithms proposed by Burgard (2013). The EBP and CEP point estimates and the corresponding estimates of the MSEs (produced by parametric bootstrap under the assumptions of the GLMM) did not differ substantially. The similarity between the EBP and the CEP estimates has been also reported by Burgard (2013). Given that the CEP methodology is the one that is used in practice, and because of the lack of substantial differences between the CEP and the EBP estimates, we report only the CEP results in this paper. The EBP estimates are available from the authors on request.

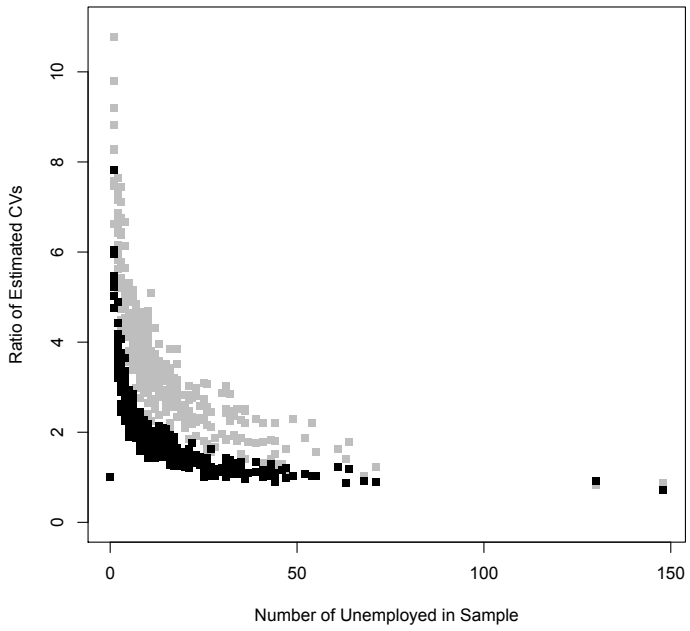

Figure 5: Ratio of estimated coefficients of variation of direct estimates to M-quantile (gray) and CEP (black) estimates of total number of unemployed people for each UALAD.

7 Model-based simulations

The validity of model-based inference depends on the validity of the assumed model. The preceding analyses of the UKLFS data are sample-specific, which makes generalisation difficult. In this Section we empirically evaluate the properties of small area predictors and corresponding MSE estimators. In particular, we use Monte Carlo simulation to carry out a sensitivity analysis of departures from GLMM model assumptions. Our simulations are model-based, in

the sense that population data are first generated under a model assumption or scenario, with a sample then selected from each simulated population. Estimates of small area proportions and corresponding MSEs are computed using the data from these samples.

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-type 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 a 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 REB bootstrap methods described in Section 5.4. Note that the logistic M-quantile linear regression fit underpinning the M-quantile and Expectile predictors was obtained by using functions 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). In addition, we generated data corresponding to a combined misclassification and measurement error scenario, denoted (M). In this scenario, a random 1% of the sample 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 CEP and the direct estimator were used to estimate the small area proportions $\bar{y}_d,\ d = 1, \ldots, D$.

The performances of these 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 4, 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 predictors under the (0) scenario, while under the (M) scenario the M-quantile predictor appears to be superior.

Table 4: Model-based simulation results: Performance of predictors of small area proportions. The true small area proportions range between 0.4 and 0.9.

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

In order to evaluate the performance of the MSE estimators for the M-quantile predictor ($c = 1.345$) proposed in Section 5.4 we used the data generated for the scenario with $D = 50$ and sample sizes $n_d = 10$ and $n_d = 20$. For providing some empirical evidence for the consistency of the proposed MSE estimators we also simulated data for the larger sample size $n_d = 30$. Again, $T = 1,000$ Monte-Carlo populations were generated and for each generated population a simple random sample without replacement of size was drawn from each area d, which was then used to calculate the M-quantile predictor and its linearisation MSE estimator (19) and the REB bootstrap MSE estimator $mse^{REB}(\hat{\bar{y}}_d^{MQ})$ based on 100 bootstrap iterations. The performance of these MSE estimators for each scenario is presented in Table 5 where we show the medians of the area-specific Bias and RMSE ($\times 1000$). We also show the medians of the empirical coverage rates for nominal 95% confidence intervals (CR95) 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 the REB bootstrap these intervals were based on the 2.5 and the 97.5 percentiles of the corresponding bootstrap distribution. Examination of the results in Table 5 shows that both MSE estimation methods tend to be biased low, but all generate nominal 95 per cent confidence intervals with acceptable coverage. Increasing the domain-specific sample sizes results in lower Bias and RMSE for both the analytic and the bootstrap MSE estimators. However, the REB bootstrap estimator exhibits smaller bias and better stability than the linearisation-based estimator (19). We therefore recommend that the REB bootstrap MSE estimator be used in applications.

8 Final remarks

Small area prediction for binary outcomes is an important and challenging problem. In this paper we propose a new approach to this problem based on an extension of M-quantile regression to binary data. By construction, the resulting M-quantile predictor is outlier robust. The benefits of the approach are illustrated by applying it to estimation of unemployment for local

Table 5: Model-based simulation results: Performance of MSE estimators.

	Median values of Bias		*Median values of RMSE* ($\times 1000$)		*Median values of CR95 (%)*	
			$n_d = 10$			
Estimator/Scenario	(0)	(M)	(0)	(M)	(0)	(M)
$mse^A(\hat{\bar{y}}_d^{MQ})$	-0.0027	-0.0031	0.1198	0.1105	95	94
$mse^{REB}(\hat{\bar{y}}_d^{MQ})$	-0.0004	-0.0006	0.0532	0.0433	95	95
			$n_d = 20$			
$mse^A(\hat{\bar{y}}_d^{MQ})$	-0.0005	-0.0019	0.1130	0.0995	95	95
$mse^{REB}(\hat{\bar{y}}_d^{MQ})$	-0.0001	-0.0006	0.0286	0.0294	95	95
			$n_d = 30$			
$mse^A(\hat{\bar{y}}_d^{MQ})$	0.0002	0.0011	0.0994	0.0893	95	95
$mse^{REB}(\hat{\bar{y}}_d^{MQ})$	-0.0000	-0.0002	0.0146	0.0172	95	95

authorities in the UK. The results indicate that the proposed methodology leads to estimates that are consistent and more efficient than direct estimates and are comparable to alternative model-based estimates. We further present two approaches to estimating the MSE of the M-quantile predictor, one based on a linearisation approach and the other based on application of REB bootstrap. The MSE estimators provide acceptable coverage performance in our simulations, with the REB bootstrap being perhaps preferable because of its stability and simplicity.

The present paper has focused on an alternative methodology to the one currently used by ONS in the UK. Molina et al. (2007) and López-Vizcaíno et al. (2013, 2014) propose the use of multinomial mixed models for multi-category outcomes with applications to small area estimation of labour force activity. The papers by López-Vizcaíno et al. (2013, 2014) further allows for category-specific random effects, which seems appropriate in practice. An obvious extension of the development set out in this paper is M-quantile modelling of multi-category outcomes. This is an area of current research.

Acknowledgment

The authors are grateful to the Editor, Associate Editor and two referees for their comments and suggestions. These led to a revised version of the article that represented a considerable improvement on the original. This work was partially supported by the PRIN 2012 project Household wealth and youth unemployment: new survey methods to meet current challenges.

Supplementary Materials

Supplementary materials are available online at .....

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

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

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.

REB bootstrap procedure

The steps in the REB bootstrap are as follows.

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

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 $h(d) = srswr(\{1,\ldots,D\},1)$.

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

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

4. Calculate the bootstrap population parameters $\bar{y}_d^*$, $d = 1,\ldots,D$.

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 $\hat{\bar{y}}_d^{MQ*},\ d = 1, \ldots, D$.

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.

7. The REB bootstrap estimator of the MSE of $\hat{\bar{y}}_d^{MQ}$ is

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

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.

Berkhof, J. and T. Snijders (2001). Variance component testing in multilevel models. *Journal of Educational and Behavioral Statistics 26*, 133–152.

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

Brown, G., R. Chambers, P. Heady, and D. Heasman (2001). Evaluation of small area estimation methods - an application to unemployment estimates from the uk lfs. *Proceedings of Statistics Canada Symposium 2001. Achieving Data Quality in a Statistical Agency: A Methodological Perspective.*.

Burgard, J. (2013). *Evaluation of Small Area Techniques for Applications in Official Statistics*. Phd-thesis, Universität Trier, 2013.

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 (2014a). Outlier robust small area estimation. *Journal of the Royal Statistical Society. Series B 76*, 47–69.

Chambers, R., E. Dreassi, and N. Salvati (2014b). Disease mapping via negative binomial m-quantile regression. *Statistics in Medicine 33*, 4805–4824.

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.

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.

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.

Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons.

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.

López-Vizcaíno, E., M. Lombardía, and D. Morales (2013). Multinomial-based small area estimation of labour force indicators. *Statistical Modelling 13*, 153–178.

López-Vizcaíno, E., M. Lombardía, and D. Morales (2014). Small area estimation of labour force indicators under a multinomial model with correlated time and area effects. *Journal of the Royal Statistical Society, Series A*, DOI: 10.1111/rssa.12085.

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.

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.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models*. London: Chapman & Hall.

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

Molina, I., A. Saei, and M. Lombardía (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., 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.

Self, S. G. and K. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association 82*, 605–10.

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 (2014). Robust small area prediction for counts. *Statistical Methods in Medical Research*, doi:10.1177/0962280214520731.

Ugarte, M., T. Goicoa, A. Militino, and M. Sagaseta-López (2009). Estimating unemployment in very small areas. *SORT 33*, 49–70.