Noname manuscript No.
(will be inserted by the editor)

Outlier Robust Small Domain Estimation Via Bias Correction And Robust Bootstrapping

Received: date / Accepted: date

Abstract Several methods have been devised to mitigate the effects of outlier values on survey estimates. If outliers are a concern for estimation of population quantities, it is even more necessary to pay attention to them in a small area estimation (SAE) context, where sample size is usually very small and the estimation in often model based. In this paper we set two goals: The first is to review recent developments in outlier robust SAE. In particular, we focus on the use of partial bias corrections when outlier robust fitted values under a working model generate biased predictions from sample data containing representative outliers. Then we propose an outlier robust bootstrap MSE estimator for M-quantile based small area predictors which considers a bounded-block-bootstrap approach. We illustrate these methods through model based and design based simulations and in the context of a particular survey data set that has many of the outlier characteristics that are observed in business surveys.

Keywords M-quantile regression · Small Area Estimation · Robust Estimation · Survey Sampling Theory · Resampling Methods

1 Introduction

1.1 Outliers in Business Survey Data

Sample outliers are a common feature of business and income surveys. They are usually identified and removed or corrected before estimation. This is non-controversial if they are non-representative outliers (Chambers, 1986). Such values correspond to data errors or are unique to the particular population element involved, and can be dealt with during the survey data editing process. However, in some cases the values associated with sample outliers are actually correct. It is just that they are quite unlike anything one might expect under the often implicit working model one has for these values. Chambers (1986) refers to these outliers as representative, i.e., there is no reason to believe that there are no more such outliers in the non-sampled population values and as a consequence they are relevant to inference about population parameters. In this paper we assume that all outlying values in the sample data correspond to representative values.

A common approach in this situation ensures that the sample outliers have minimal or no effect on the population estimates derived from the sample data. There are a variety of ways to do this, but one that is becoming increasingly popular is to fit the working model to the sample data using outlier robust estimation methods, then use this robust fit to define the population estimates, typically via prediction of the non-sampled population values based on this robust fit. As Chambers et al. (2014) note, this corresponds to 'projecting' the robust fit onto the non-sampled part of the population. Or, since the robust fit is essentially a working model fit for those sample values that are not outliers, effectively assuming that there are no outliers (relative to the working model) in the non-sampled data. This is clearly not correct if there are representative sample outliers, and implies that we can expect this approach to be biased. In such cases, Chambers (1986) suggests the addition of an extra term to the robust projective estimator that partially corrects for its bias under the working model. Since outlier robust prediction is essentially a bias-variance trade-off, the purpose of this term is not to completely remove the bias associated with the robust projective fit, since that would lead to a highly non-robust predictor. This term allows the sample outliers to contribute to the prediction of the non-sampled population values in a controlled fashion, typically by not giving them full weight in the bias correction term.

The outlier problem can be magnified considerably if, rather than population estimation, the target is estimation for small domains (Chambers and Tzavidis, 2006; Sinha and Rao, 2009; Giusti et al., 2012; Jiongo et al., 2013; Chambers et al., 2014). Clearly a single outlier in a large population sample has much less relative impact than the same outlier in a much smaller domain sample. However, estimation for such domains, particularly when these correspond to small geographic areas, or, in the case of business surveys, fine level industry classes, is now increasingly important for business surveys. In this context, it seems appropriate to explore domain level implementation of the strategy of outlier robust model fitting followed by partial bias correction that was described in the previous paragraph. Chambers et al. (2014) refer to such a strategy as being predictive while, as already noted above, one that just replaces working model parameters by robust estimates is referred to as being projective. They show that robust-predictive estimators are less biased and can be more efficient than the robust-projective estimators, e.g. the Robust EBLUP of Sinha and Rao (2009) and M-quantile predictor of Chambers and Tzavidis (2006), that have been suggested in the literature. However, the bias correction used by robust-predictive estimators comes at a cost of higher variability, and so use of these estimators is recommended when model diagnostics suggest that the sample data contain values that represent significant departures from the first order properties of the working model.

Estimation of the mean squared error (MSE) of an outlier robust small area predictor is difficult. In the absence of a suitable analytical approach for estimating the MSE of the robust predictor developed in their paper, Sinha and Rao (2009) propose a MSE estimator based on a parametric bootstrap. However, this approach has limitations since it essentially assumes that all sample outliers are non-representative. Chambers et al. (2014) propose an analytic linearization-based MSE estimator for an outlier robust predictor of a small area mean based on first-order approximations to the variances of solutions of robust estimating equations. The authors show that this approach offers promising results when used in conjunction with robust-predictive estimation methods. However, it is also complex to calculate, and can be unstable in the large skewed samples characteristic of business surveys.

The purpose of this paper is two-fold. The first is to review recent developments in outlier robust small domain estimation, or, as it is more usually referred to, small area estimation (SAE). In particular, we focus on the use of partial bias corrections when outlier robust fitted values under a working model generate biased predictions from sample data containing representative outliers. The second is to propose an outlier robust bootstrap MSE estimator for M-quantile based small area predictors (Chambers et al., 2014), extending the block-bootstrap approach of Chambers and Chandra (2013). We illustrate our ideas in the context of a particular survey data set (the ABI data, see Section 1.2) that has many of the outlier characteristics that are observed in business surveys. Sections 2 and 3 of the paper therefore review recent developments in outlier robust SAE, while Section 4 describes the proposed robust bootstrap-based MSE estimator for M-quantile predictors. In Section 5 we then apply the proposed methods to the ABI data. Section 6 concludes the paper with a discussion. Model-based and Design-based simulations that replicates the ABI data structure are carried out in order to compare the performance of the predictors and corresponding MSE estimators discussed in Sections 2 to 4. In order to ensure that this paper remains a reasonable length, the results of these simulations are provided in the Supplementary Materials.

1.2 An Example: The ABI Sample

The Annual Business Inquiry (ABI) dataset was created by the UK Office for National Statistics in order to provide a realistic demonstration of outlier-prone business data for use in evaluating the automatic outlier detection editing methods that were developed as part of the Euredit project (Office for National Statistics, 2000). Here we do not pursue that objective, but instead view the $n = 6,099$ cases making up this data set as illustrative of business survey data containing outliers, and focus on domain estimation based on the anonymised industrial classifier CLASS that is part of the ABI data set. This variable has 28 categories, with sample sizes that range from 9 to 1553, and where 8 categories have sample sizes less than 50. The main auxiliary variable in the ABI data set is the value of turnover as recorded in the business population register, denoted TURNREG, while the main output variables are actual total turnover (TURNOVER), total employment costs (EMPTOTC), total purchases of goods and services (PURTOT), total taxes paid (TAXTOT) and total number of employees (EMPLOY).

We focus on TAXTOT and restrict attention to those business with TAXTOT > 0 and with TURNREG strictly between 0 and 100,000. There are 5,554 such cases, corresponding to an estimated population (defined by the sum of the sample weights of these businesses) of 193,164 businesses. Table 1 shows the samples and estimated populations counts for the 28 CLASS groups. Our objective is estimation of the population mean of TAXTOT for each level of CLASS.

The relationship between TURNREG and TAXTOT for the ABI businesses is shown in Figure 1,vand indicates a level of heterogeneity that is not unusual for a business survey data set. As a consequence, it may not be appropriate to use a linear mixed model to estimate population quantities relating to TAXTOT for the ABI population. One of the assumptions in such a model is that within-

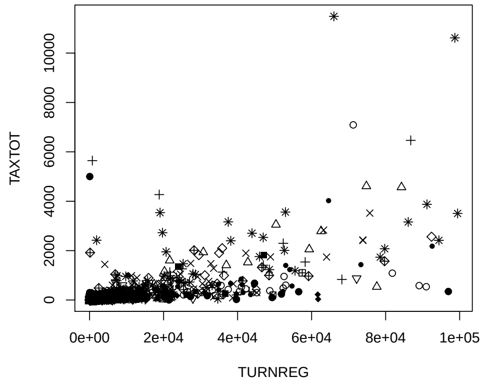
CLASS	11	21	22	23	24	25	26
n	226	134	293	1489	183	201	62
N	4896	3253	7076	44247	5069	4711	1548
CLASS	27	28	31	32	33	34	35
n	586	323	27	103	181	153	143
N	17316	10080	1533	5291	10087	10554	6745
CLASS	36	37	41	42	43	51	52
n	83	63	35	66	131	266	409
N	4948	2547	1106	1693	3773	14151	20807
CLASS	61	62	63	71	72	73	74
n	269	47	18	7	17	17	22
N	6693	715	200	614	1461	796	1254

Table 1 CLASS level sample and population counts for the restricted ABI sample.

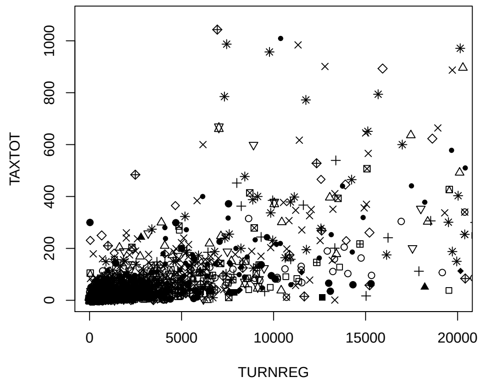
domain variances are homogeneous. This assumption may not be appropriate when the heterogeneity in the data is high, particularly when estimation of domain-specific variances is of interest, with the aim of predicting domain-specific random effects. In Figure 2 we show four examples of the CLASS level relationships between TAXTOT and TURNREG, noting the presence of a large number of representative outlier values. Sections 2 to 4 discuss methods for controlling their impact on estimation of the CLASS means of TAXTOT, and we apply these methods to the ABI data set in Section 5.

2 Outlier Robust SAE

We assume that the target population (and sample) can be divided into D domains or small areas, indexed by i below, with population units indexed by j. The overall population (sample) size is denoted by $N(n)$, with $N_i(n_i)$ population (sample) units in area i. We denote the population (sample) units in area i by $U_i(s_i)$. The variable of interest is denoted y, with values y_{ij} and is assumed to be measured (at least conceptually) on a continuous scale. The target parameters are the overall population mean, denoted $\overline{y}_U$, as well as the small area means $\overline{y}_i$. Note that 'small' in this context usually means



(a)



(b)

Fig. 1 Plot of TAXTOT vs. TURNREG for ABI sample. Plot on left is for all 5,554 sample businesses, while plot on right is for the 5,363 sample businesses with TURNREG $<$ 20,000 and TAXTOT $<$ 1,000. Different CLASS groups are indicated by different plotting symbols.

that the sample size n_i in area i is too small to allow standard domain estimation with sufficient accuracy, and so models are typically used to 'borrow strength' from other areas for estimating $\overline{y}_i$. See Rao and Molina (2015) for an overview of what this concept implies. We just note here that it includes two key assumptions. The first is that our working model for the conditional

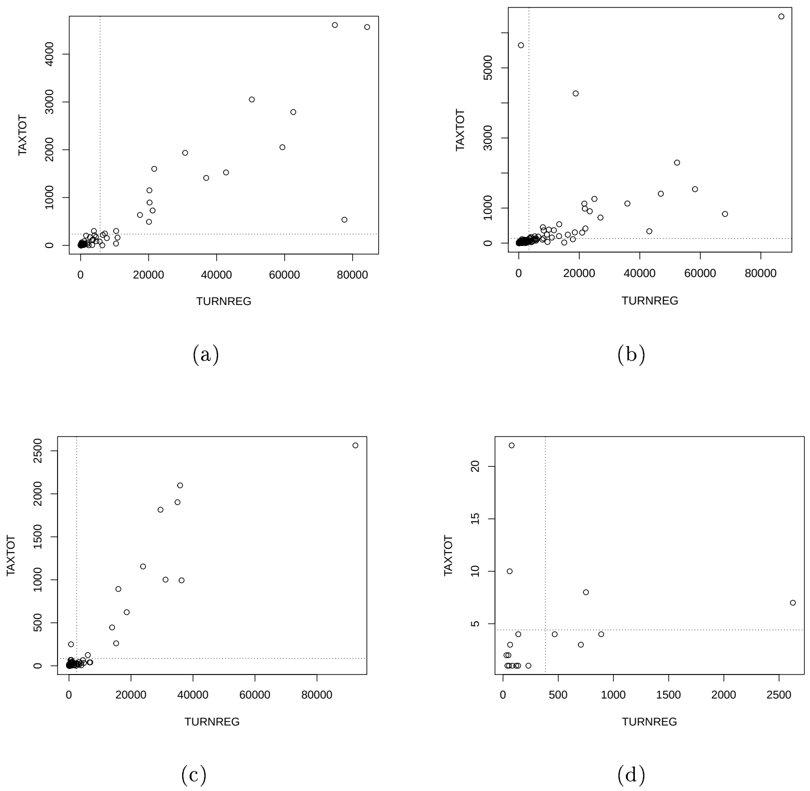


Fig. 2 Plot of TAXTOT vs. TURNREG for four CLASS groups. Dotted lines are mean values.

distribution of y_{ij} given the value $\boldsymbol{x}_{ij}$ of suitable covariates whose values are assumed known for all population units adequately characterises the between and within area variability of y_{ij}. The second is that our method of sampling is non-informative for this conditional distribution. That is, we can validly assume that the non-sampled population units have the same conditional distribution as the sampled population units.

Under this set up it is not difficult to see that a representative sample outlier can call into question the overall adequacy of the working model, insofar as it clearly implies that there are sample values y_{ij} that are not well fitted by it.

Furthermore, since these sample outliers are correct population values, there is no indication that similar values might not also exist in the non-sampled values of y. It is then a small step to realising that, given the much smaller sample sizes involved, such a population outlier in area i can have a much greater impact on estimation for the mean $\overline{y}_i$ of the area than it will have on estimation for the overall population mean $\overline{y}_U$. Use of outlier robust SAE methods with business survey data therefore seems a sensible strategy.

2.1 Linear Random Effects Modeling for SAE

Most business survey estimation methods are based on the implicit use of working models that are linear in the covariates. Consequently, we restrict ourselves in what follows to SAE models that are also linear. In this section we summarise the standard random intercepts unit level working model used in SAE. This is the linear mixed model

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + u_i + e_{ij}, \tag{1}$$

where u_i is an area (level 2) random effect and e_{ij} is an individual (level 1) random effect. The area effects are assumed to all have zero mean and the same variance, and are distributed independently of the individual effects, which are also assumed to have zero mean and the same variance. We write this as $\boldsymbol{u} = (u_i) \sim iid(0, \sigma_u^2)$ and $\boldsymbol{e} = (e_{ij}) \sim iid(0, \sigma_e^2)$, where $\boldsymbol{u}$ and $\boldsymbol{e}$ are independent. The parameters σ_u^2 and σ_e^2 are usually referred to as the variance components of the model and the covariance matrix of the vector $\boldsymbol{y}_i = (y_{ij})$ of population values of y in area i is then $\boldsymbol{V}_i = \sigma_e^2 \boldsymbol{I}_{N_i} + \sigma_u^2 \boldsymbol{1}_{N_i} \boldsymbol{1}_{N_i}^T$, where $\boldsymbol{I}_{N_i}$ is the identity matrix of order N_i and $\boldsymbol{1}_{N_i}$ is a vector of ones of the same order.

Methods for fitting a model (1) are now widely available. These typically assume random effects are normally distributed and lead to an estimate $\hat{\boldsymbol{\beta}}$ of the regression parameters (typically referred to as fixed effects) as well as predicted values $\hat{u}_i$ for the area i random effects (which depends on $\hat{\boldsymbol{\beta}}$ as well as on the estimates of the model variance components). The predicted value for y_{ij} for non-sampled population unit j in area i is then

$$\widehat{y}_{ij} = \boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}} + \widehat{u}_i, \tag{2}$$

and Empirical Best Linear Unbiased Predictor of small area mean for area i, $\overline{y}_i$, is (Battese et al., 1988)

$$EBLUP_i = N_i^{-1}\left\{ \sum_{j\in S_i} y_{ij} + \sum_{j\in U_i \setminus s_i} \left(\boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}} + \widehat{u}_i \right) \right\} = N_i^{-1}\left\{ \sum_{j\in s_i} y_{ij} + \sum_{j\in U_i\setminus s_i} \widehat{y}_{ij} \right\}, \tag{3}$$

where $U_i \setminus s_i$ indicates the non-sample units in small area i. A linearisation-based estimator of the MSE of (3) was developed by Prasad and Rao (1990). See also Rao and Molina (2015, Section 7.2, pp.179-186). This estimator is known to work well when the model (1) holds. It is denoted by PR in what follows. Alternatively, a bootstrap method can be used for MSE estimation. For details see González-Manteiga et al. (2007) and Hall and Maiti (2006).

2.2 Outlier Robust EBLUP

Sinha and Rao (2009) tackle the problem of how to deal with outliers in SAE under the working model (1) by replacing $\widehat{\boldsymbol{\beta}}$ and $\widehat{u}_i$ in (2) with outlier robust versions $\widehat{\boldsymbol{\beta}}^{\Psi}$ and $\widehat{u}_i^{\Psi}$, leading to the outlier robust predictor

$$\widehat{y}_{ij}^{\Psi} = \boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}}^{\Psi} + \widehat{u}_i^{\Psi}. \tag{4}$$

The outlier robust estimates of the fixed and random effects in (1) are defined by combining a modified version of the robust estimating equations for $\boldsymbol{\beta}$,

$$\sum_i \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{U}_i^{1/2} \boldsymbol{\Psi}(\boldsymbol{U}_i^{-1/2} \boldsymbol{r}_i) = \boldsymbol{0}, \tag{5}$$

with the robust Maximum Likelihood Estimation Method 2 (MLE2) estimating equations for the variance components $\boldsymbol{\theta} = (\theta_l) = (\sigma_u^2, \sigma_e^2)$ developed by Sinha and Rao (2009),

$$\sum_i \left\{ \boldsymbol{\Psi}^T (\boldsymbol{U}^{-1/2} \boldsymbol{r}_i) \boldsymbol{U}^{1/2} \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V} i}{\partial \theta_l} \boldsymbol{V}_i^{-1} \boldsymbol{U}^{1/2} \boldsymbol{\Psi}(\boldsymbol{U}^{-1/2} \boldsymbol{r}_i) - k_2 tr \left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \theta_l} \right) \right\} = 0 \tag{6}$$

leading to a robust EBLUP for $\overline{y}_i$,

$$SR_i = N_i^{-1}\left\{\sum_{j\in s_i} y_{ij} + \sum_{j\in U_i\setminus s_i} \widehat{y}_{ij}^{\boldsymbol{\Psi}}\right\}, \tag{7}$$

where the vector of random area-specific effect $\boldsymbol{u} = (u_i)$ is estimated following Fellner (1986). Here $\boldsymbol{\Psi}$ is a bounded, skew-symmetric influence function that controls the influence of level 2 and level 1 outliers on the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{X}_i^T\boldsymbol{\beta}$, $\boldsymbol{U} = diag(\boldsymbol{V})$ and $k_2 = E[\boldsymbol{\Psi}^2(Z)]$ where $Z \sim N(0,1)$. A popular choice for $\boldsymbol{\Psi}$ is the Huber type influence function with tuning constant $c = 1.345$ in (5) and $c = 2$ in (6). As an aside, we note that use of (6) to estimate the variance components associated with the 'non-outliers' in a data set leads to downward bias when outliers are present, an inevitable compromise when trying to estimate variances while simultaneously restricting the influence of large positive residuals on the variance estimates.

Sinha and Rao (2009) proposed a parametric bootstrap for estimating the MSE of estimator (7), but also noted that this model-based bootstrap technique has a limitation, in the sense that it depends on the working assumption about the distributions of the area-specific effects and the random unit errors. A linearisation-based MSE estimator for (7) was developed by Chambers et al. (2014). This combines the approach of Booth and Hobert (1998) with the linearisation approach to estimation of prediction variance of Street et al. (1988), to obtain the approximation

$$mse(SR_i) = Var(SR_i) + \{Bias(SR_i)\}^2,$$

where $Var(SR_i) = h_{1i}(\widetilde{\boldsymbol{\delta}}) + h_{2i}(\widetilde{\boldsymbol{\delta}}) + h_{3i}(\widetilde{\boldsymbol{\delta}})$ with $\widetilde{\boldsymbol{\delta}} = (\widetilde{\boldsymbol{\beta}}^{\boldsymbol{\Psi}T}, \widetilde{\boldsymbol{u}}^{\boldsymbol{\Psi}T})^T$ and $Bias(SR_i)$ is the conditional (i.e., given the value of the area i effect) bias of SR_i as a predictor of $\overline{y}_i$. Here $\widetilde{\boldsymbol{\delta}} = (\widetilde{\boldsymbol{\beta}}^{\boldsymbol{\Psi}T}, \widetilde{\boldsymbol{u}}^{\boldsymbol{\Psi}T})^T$ are estimated values of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ that are obtained by solving (5) when variance components are known and

- $h_{1i}(\widetilde{\boldsymbol{\delta}}) = (1 - n_i N_i^{-1})^2(\overline{\boldsymbol{x}}_{ri}^T \;\; \overline{\boldsymbol{z}}_{ri}^T)\widehat{\boldsymbol{V}}_{\boldsymbol{u}}(\widetilde{\boldsymbol{\delta}})(\overline{\boldsymbol{x}}_{ri}^T \;\; \overline{\boldsymbol{z}}_{ri}^T)^T$ is the variance contribution due to robust estimation of the fixed and random effects in the linear mixed model, where $\overline{\boldsymbol{x}}_{ri}$ and $\overline{\boldsymbol{z}}_{ri}$ denote the vectors of average values of $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ respectively for the $N_i - n_i$ non sampled units in area i, with $\boldsymbol{z}_{ij}$ denoting the vector of dimension D with value one in position i and zeros elsewhere and $\widehat{\boldsymbol{V}}_u(\widetilde{\boldsymbol{\delta}})$ is the first order approximation to the conditional covariance matrix $Var(\widetilde{\boldsymbol{\delta}}|\boldsymbol{u})$ defined in Chambers et al. (2014).

- $h_{2i}(\widetilde{\boldsymbol{\delta}}) = (1 - n_i N_i^{-1})^2 \boldsymbol{V_u}(\overline{e}_{ri})$ is the variance contribution associated with prediction of $\overline{y}_{ri}$, the area i non-sample average value of y, i.e.

$$\boldsymbol{V_u}(\overline{e}_{ri}) = Var(\overline{y}_{ri}|\boldsymbol{u}) = Var\left((N_i - n_i)^{-1} \sum_{j \in U_i \setminus s_i} (y_{ij} - \boldsymbol{x}_{ij}^T \widetilde{\boldsymbol{\beta}}^{\boldsymbol{\Psi}} - \widetilde{u}_i^{\boldsymbol{\Psi}})|\boldsymbol{u} \right).$$

- $h_{3i}(\widetilde{\boldsymbol{\delta}}) = \left(N_i^{-1} \sum_{j \in U_i \setminus s_i} \boldsymbol{z}_{ij}^T \right) \boldsymbol{\Upsilon} \left(N_i^{-1} \sum_{j \in U_i \setminus s_i} \boldsymbol{z}_{ij}^T \right)^T$ is the variance contribution due to robust estimation of the variance components of the linear mixed model, with

$$\boldsymbol{\Upsilon} = \sum_{a=1}^{2} \sum_{b=1}^{2} \left\{ \left(\partial_{\boldsymbol{\theta}_a} \boldsymbol{G} \right) \Big[\sum_{i=1}^{D} \sum_{k=1}^{D} \left\{ \widetilde{u}_i^{\boldsymbol{\Psi}} \widetilde{u}_k^{\boldsymbol{\Psi}} + \sigma_e^2 \boldsymbol{I}(i = k) \right\} \Big] (\partial_{\boldsymbol{\theta}_b} \boldsymbol{G})^T \boldsymbol{V}_{\boldsymbol{u}ab}(\widehat{\boldsymbol{\theta}}) \right\},$$

where $\boldsymbol{V_u}(\widehat{\boldsymbol{\theta}}) = [\boldsymbol{V}_{\boldsymbol{u}ab}(\widehat{\boldsymbol{\theta}})] = Var(\widehat{\boldsymbol{\theta}}|\boldsymbol{u})$ and $\boldsymbol{G}$ is an $D \times n$ matrix

$$\boldsymbol{G} = \left(\boldsymbol{Z}^T (\sigma_e^2 \boldsymbol{I}_n)^{-1/2} \boldsymbol{W}_2 (\sigma_e^2 \boldsymbol{I}_n)^{-1/2} \boldsymbol{Z} + (\sigma_u^2 \boldsymbol{I}_m)^{-1/2} \boldsymbol{W}_3 (\sigma_u^2 \boldsymbol{I}_m)^{-1/2} \right)^{-1} \left(\boldsymbol{Z}^T (\sigma_e^2 \boldsymbol{I}_n)^{-1/2} \boldsymbol{W}_2 (\sigma_e^2 \boldsymbol{I}_n)^{-1/2} \right),$$

with $\boldsymbol{W}_2$ an $n \times n$ diagonal matrix of weights with $j-$th component

$$w_{2ij} = \left[(\sigma_e)^{-1} (y_{ij} - \boldsymbol{x}_{ij}^T \widetilde{\boldsymbol{\beta}}^{\boldsymbol{\Psi}} - \widetilde{u}_i^{\boldsymbol{\Psi}}) \right]^{-1} \boldsymbol{\Psi} \left\{ (\sigma_e)^{-1} (y_{ij} - \boldsymbol{x}_{ij}^T \widetilde{\boldsymbol{\beta}}^{\boldsymbol{\Psi}} - \widetilde{u}_i^{\boldsymbol{\Psi}}) \right\},$$

and $\boldsymbol{W}_3$ an $D \times D$ diagonal matrix of weights with $i-$th component

$$w_{3ij} = \left[(\boldsymbol{\sigma}_u)^{-1} \widetilde{u}_i^{\boldsymbol{\Psi}} \right]^{-1} \boldsymbol{\Psi} \left\{ (\boldsymbol{\sigma}_u)^{-1} \widetilde{u}_i^{\boldsymbol{\Psi}} \right\}.$$

See Chambers et al. (2014) for further details. This MSE approximation can be estimated by plugging in robust estimates of the model parameters and the area effects. We denote the resulting estimator by CST in what follows, and note that there are two ways in which the marginal variance $V_{\boldsymbol{u}}(\overline{e}_{ri})$ can be estimated:

– conditionally, via

$$\hat{V}_{\boldsymbol{u}}(\bar{e}_{ri}) = (N_i - n_i)^{-1}(n_i - p)^{-1} \sum_{j \in s_i} (y_{ij} - \boldsymbol{x}_{ij}^T \hat{\boldsymbol{\beta}}^{\boldsymbol{\Psi}} - \hat{u}_i^{\boldsymbol{\Psi}})^2,$$

– unconditionally, via

$$\hat{V}_{\boldsymbol{u}}(\bar{e}_{ri}) = (N_i - n_i)^{-1}(n - p)^{-1} \sum_k \sum_{j \in s_k} (y_{kj} - \boldsymbol{x}_{kj}^T \hat{\boldsymbol{\beta}}^{\boldsymbol{\Psi}} - \hat{u}_k^{\boldsymbol{\Psi}})^2,$$

where p is the number of covariates. The version of CST that is based on the unconditional (pooled) estimator above is usually more stable, and hence preferred.

2.3 Predictive Outlier Robust SAE

Since the outlier robust predictor SR_i minimises the influence of sample outliers on prediction, it will be biased when representative sample outliers are drawn from a distribution with a mean structure that differs from that implied by the working model. As a consequence, Chambers et al. (2014) refer to a predictor like SR_i as 'projective', insofar as it projects working model behaviour on to all non-sample population units. On the other hand, the outlier robust predictor of Chambers (1986) modifies a projective predictor to allow for representative outliers to contribute, albeit in a controlled way. Chambers et al. (2014) refer to this approach as 'predictive', and use it to develop an alternative to SR_i. This alternative predictor corresponds to the mean functional defined by the Welsh and Ronchetti (1998) outlier robust predictor of the finite population distribution of y in area i. This is

$$\hat{\boldsymbol{F}}_i^{\boldsymbol{\Psi\Phi}}(t) = N_i^{-1}\Big[\sum_{j \in s_i} \boldsymbol{I}(y_{ij} \leqslant t) + \sum_{j \in U_i \setminus s_i} \hat{\boldsymbol{P}}_{ij}^{\boldsymbol{\Psi\Phi}}(t)\Big], \tag{8}$$

where $\hat{\boldsymbol{P}}_{ij}^{\boldsymbol{\Psi\Phi}}(t) = n_i^{-1} \sum_{k \in s_i} \boldsymbol{I}\big(\hat{y}_{ij}^{\boldsymbol{\Psi}} + \hat{\boldsymbol{\omega}}_i^{\boldsymbol{\Psi}} \boldsymbol{\Phi}\{r_{ik}^{\boldsymbol{\Psi}}/\hat{\boldsymbol{\omega}}_i^{\boldsymbol{\Psi}}\} \leqslant t\big)$, $r_{ik}^{\boldsymbol{\Psi}} = y_{ik} - \hat{y}_{ik}^{\boldsymbol{\Psi}}$,$\hat{\boldsymbol{\omega}}_i^{\boldsymbol{\Psi}}$ is a robust estimate of the scale of the $r_{ik}^{\boldsymbol{\Psi}}$; $k \in s_i$ and $\boldsymbol{\Phi}$ is another bounded influence function that is not as restrictive as $\boldsymbol{\Psi}$, i.e. $|\boldsymbol{\Phi}| \geqslant |\boldsymbol{\Psi}|$. Typically, both $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are Huber-type influence functions defined by tuning constants

k and c respectively, and we have $k \leqslant c$. In what follows we put $k = 1.345$ and $c = 3$. The proposed predictor of $\overline{y}_i$ is then obtained by integrating the identity function with respect to (8), which leads to

$$SRB_i = \int t d\hat{F}_i^{\boldsymbol{\Psi}\boldsymbol{\Phi}}(t) = SR_i + \left(1 - n_i N_i^{-1}\right)\hat{B}_i, \tag{9}$$

where $\hat{B}_i = n_i^{-1} \sum_{j \in s_i} \hat{\boldsymbol{\omega}}_i^{\boldsymbol{\Psi}} \boldsymbol{\Phi}(r_{ik}^{\boldsymbol{\Psi}} / \hat{\boldsymbol{\omega}}_i^{\boldsymbol{\Psi}})$ is a robust bias correction that allows representative outliers to have a larger effect on the predicted value of $\overline{y}_i$ than allowed by SR_i. MSE estimation for (9) can also be carried out using CST, which in this case is based on the approximation

$$mse(SRB_i) = h_{1i}^B(\tilde{\boldsymbol{\delta}}) + h_{2i}(\tilde{\boldsymbol{\delta}}) + h_{3i}^B(\tilde{\boldsymbol{\delta}}) + h_{4i}^B(\tilde{\boldsymbol{\delta}}),$$

where h_{2i} was defined earlier in Section 2.2 and, under the working model (1),

- $h_{1i}^B(\tilde{\boldsymbol{\delta}}) = (1 - n_i N_i^{-1})^2 \begin{pmatrix} \overline{\boldsymbol{x}}_{ri} - \overline{\boldsymbol{x}}_{si} \\ \mathbf{0}_D \end{pmatrix}^T \hat{\boldsymbol{V}}_{\boldsymbol{u}}(\tilde{\boldsymbol{\delta}}) \begin{pmatrix} \overline{\boldsymbol{x}}_{ri} - \overline{\boldsymbol{x}}_{si} \\ \mathbf{0}_D \end{pmatrix}$, where $\hat{\boldsymbol{V}}_{\boldsymbol{u}}(\tilde{\boldsymbol{\delta}})$ was defined earlier in Section 2.2, and $\overline{\boldsymbol{x}}_{si}$ is the simple average of $\boldsymbol{x}_{ij}$ in area i;
- $h_{3i}^B(\tilde{\boldsymbol{\delta}}) = (1 - n_i N_i^{-1})^2 n_i^{-1} (n_i - p - 1)^{-1} \sum_{j \in s_i} \left\{ \boldsymbol{\omega}_i^{\boldsymbol{\Psi}} \boldsymbol{\Phi} \left\{ \boldsymbol{\omega}_i^{\boldsymbol{\Psi}} \boldsymbol{\Phi} \left(\frac{y_{ik} - \boldsymbol{x}_{ij}^T \tilde{\boldsymbol{\beta}}^{\boldsymbol{\Psi}} - \tilde{u}_i^{\boldsymbol{\Psi}}}{\boldsymbol{\omega}_i^{\boldsymbol{\Psi}}} \right) \right\} \right\}^2$;
- $h_{4i}^B(\tilde{\boldsymbol{\delta}}) = (1 - n_i N_i^{-1})^2 \boldsymbol{D}_i^T \boldsymbol{\Upsilon} \boldsymbol{D}_i$, where $\boldsymbol{\Upsilon}$ was defined in Section 2.2, and

$$\boldsymbol{D}_i = \overline{\boldsymbol{z}}_{ri} - n_i^{-1} \sum_{j \in s_i} \Phi' \left[\left\{ y_{ij} - \boldsymbol{x}_{ij}^T \tilde{\boldsymbol{\beta}}^{\boldsymbol{\Psi}} - \boldsymbol{z}_{ij}^T \boldsymbol{G} (\boldsymbol{y}_s - \boldsymbol{X}_s \tilde{\boldsymbol{\beta}}^{\boldsymbol{\Psi}}) \right\} / \boldsymbol{\omega}_i^{\boldsymbol{\Psi}} \right] \boldsymbol{z}_{ij}.$$

  Here $\boldsymbol{G}$ is the same as in Section 2.2., $\boldsymbol{y}_s$ is the $n \times 1$ vector of sample values of y and $\boldsymbol{X}_s$ is the $n \times p$ matrix of sample values for the model covariates. See Chambers (1986) for details.

2.4 A More Suitable Outlier Robust Predictor

From (9), we see that the bias correction $\hat{B}_i$ in SRB_i is a direct estimate. This has the advantage that no information from representative outliers not in area i is used in calculating this bias correction. However, it also means that $\hat{B}_i$ can be variable when the area sample size n_i is small. Jiongo et al.

(2013) have proposed a more stable bias correction that allows representative outlier information to be 'shared' across areas when computing the correction for a particular area. Using an extension of the argument used to motivate the Chambers (1986) robust predictor the authors first write

$$
\begin{aligned}
EBLUP_i =& N_i^{-1} \sum_h \sum_{k\in s_h} w_{ihk} y_{hk} = \\
&= EBLUP_i + N_i^{-1} \begin{pmatrix} \sum_{j\in s_i} (w_{iij}-1) r_{ij} + (\sum_{j\in s_i} w_{iij} - N_i)\hat{u}_i \\ \sum_{h\neq i} \sum_{k\in s_h} w_{ihk} r_{hk} + \sum_{h\neq i} (\sum_{k\in s_h} w_{ihk})\hat{u}_h \end{pmatrix}
\end{aligned}
$$

where $r_{ij} = y_{ij} - \boldsymbol{x}_{ij}^T \hat{\boldsymbol{\beta}} - \hat{u}_i$ and the weights satisfy the relations

$$
w_{ihk}(\delta) = \begin{cases} D^{-1} \boldsymbol{a}_i \boldsymbol{X}_h^T \boldsymbol{C}_h^{(j)} (k \in s_h), \\ 1 + D^{-1} \boldsymbol{a}_i \boldsymbol{X}_i^T \boldsymbol{C}_i^{(j)} + (N_i - n_i)\sigma_v^2 \mathbf{1}_{n_i}^T \boldsymbol{C}_i^{(j)} (j \in s_i) \end{cases}
$$

with

$$
\boldsymbol{a}_i = \left\{ \sum_{k\in U_i \setminus s_i} \boldsymbol{x}_{ij}^T - \sigma_v^2 (N_i - n_i) \mathbf{1}_{n_i}^T \boldsymbol{V}_i^{-1}(\delta) \boldsymbol{X}_i \}\{ D^{-1} \sum_{i=1}^{D} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1}(\delta) \boldsymbol{X}_i \right\}^{-1}
$$

and $\boldsymbol{C}_i(\delta) = \boldsymbol{V}_i^{-1}(\delta)$ is a matrix satisfying $\boldsymbol{C}_i(\delta) \equiv \boldsymbol{C}_i = (\boldsymbol{C}_i^{(1)}, \ldots, \boldsymbol{C}_i^{(n_i)})$ with $\boldsymbol{C}_i^{(k)}$ corresponding to the kth column of $\boldsymbol{C}_i$ (see Jiongo et al., 2013, for further details). Replacing all non-robust quantities on the right hand side of this identity by corresponding robust values and then bounding residuals leads to the predictor

$$
\begin{aligned}
SRC_i =& SR_i + N_i^{-1} \begin{pmatrix} \sum_{j\in s_i} \boldsymbol{\Phi}_{c_1}\{(w_{iij}-1) r_{ij}^{\boldsymbol{\Psi}}\} + \boldsymbol{\Phi}_{c_2}\{(\sum_{j\in s_i} w_{iij} - N_i)\hat{u}_i^{\boldsymbol{\Psi}}\} \\ \sum_{h\neq i} \sum_{k\in s_h} \boldsymbol{\Phi}_{c_1}\{w_{ihk} r_{hk}^{\boldsymbol{\Psi}}\} + \sum_{h\neq i} \boldsymbol{\Phi}_{c_2}\{(\sum_{k\in s_h} w_{ihk})\hat{u}_h^{\boldsymbol{\Psi}}\} \end{pmatrix} \\
&= SR_i + N_i^{-1} \sum_h \sum_{j\in s_h} \boldsymbol{\Phi}_1\{(w_{ihj} - \boldsymbol{I}(j\in i)) r_{kj}^{\boldsymbol{\Psi}}\} + N_i^{-1} \sum_h \boldsymbol{\Phi}_2\{W_{ih} \hat{u}_h^{\Psi}\},
\end{aligned}
\tag{10}
$$

where $W_{ih} = (\sum_{k\in s_h} w_{ihk}) - I(h=1)N_i)$ and $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ are two other bounded influence functions. Again, these are typically Huber-type functions with tuning constants c_1 and c_2 respectively. Jiongo et al. (2013) suggest $c_1 = c\hat{\sigma}_e^{\boldsymbol{\Psi}} \times median(w_{ihj})$ and $c_2 = c\hat{\sigma}_u^{\boldsymbol{\Psi}} \times median(W_{ih})$ where c could be

3, 6 and 9. For details on this choice see the results of the simulation experiments in Jiongo et al. (2013). Observe that the robust bias correction term in SRC_i is an indirect estimate that uses information from the entire sample. Also note that since weighted residuals are bounded, this correction term is essentially protected against extreme EBLUP weights as well as against extreme residuals. Finally, we note that Jiongo et al. (2013) suggest that MSE estimation of (10) be carried out by a parametric bootstrap.

2.5 Numerical Issues with SR-based Predictors

The algorithm used to calculate the robust predictor SR_i is complex and does not always converge, as demonstrated in the ABI data analysis reported in Section 5. Furthermore, even when it does converge, it can do so quite slowly. To illustrate, in Table 2 we show, for different sample sizes of ABI data, the time to convergence (in minutes) of the R functions

- *lme* (*nlme* package in R) - used to calculate the REML fit to the linear mixed model used in $EBLUP_i$;
- *rlmer* (*robustlmm* package in R) - used to calculate the robust fit to the linear mixed model employed in SR_i;
- *QRLM* (a slightly modified version of the *rlm* code in the *mass* package of R) - used to calculate the M-quantile model fit underpinning the MQ approach described in Section 3. It is available from the authors on request.

In all cases these times were obtained using R operating under MacOS on a 2 GHz Intel i7 with 16 Gb RAM. The time to convergence obviously increases with the sample size, but looks reasonable for *lme* and *QRLM*. The *rlmer* times, on the other hand, are much greater and indicate that this function might not be practical for large sample sizes. Furthermore, although we do not show it here, these execution and convergence issues can become even more problematic when we compute the linearised mean squared error estimate CST for SRB_i and SRC_i. Finally, as mentioned earlier, we note that there are also

n	lme	rlmer	QRLM
1000	0.00052	0.46567	0.00082
2000	0.00047	0.89308	0.00065
4000	0.00067	3.10560	0.00445
6099	0.00100	6.82303	0.00220

Table 2 Convergence times (in minutes) of R functions used to fit different models to samples from the ABI data.

bias issues for the robust estimates of the variance components calculated by solving (5) and (6). This can affect choice of the tuning constants, but more importantly can impact on the bias of the linearised mean squared error estimator CST for the robust predictors SR_i and SRB_i and as well as on the validity of the parametric bootstrap-based mean squared error estimator that might be considered as an alternative.

3 M-quantile Models: A Distribution Free Alternative to Mixed Effects Models

3.1 M-quantiles Regression

The basis of the model-based approach to SAE is that population units in different areas exhibit sufficiently different distributions for the target variable y (even after conditioning on the values of the covariates $\boldsymbol{x}$) to warrant also conditioning on knowledge of which area a unit is drawn from. Sometimes, these differences are ascribed to between area differences due to the values of unknown variables. In any case, when there are relatively few areas (or domains of interest) one can usually extend the working model to include fixed area effects, thus allowing these differences to be modelled. However, when there are many areas, with small samples in each, the more common strategy is to make the area effects random and move to a mixed effects working model. However, this is not the only strategy one can adopt in this situation. Another strategy is to directly model the between area differences in the conditional distribution of y given $\boldsymbol{x}$ by associating different regions in the support for this conditional distribution with different areas. Such regions can be defined by a suitable indexing of this support, and recent SAE research has focussed on one particular indexing, corresponding to the M-quantiles of the conditional distribution of y given $\boldsymbol{x}$ across the target population. This approach has three major advantages over mixed modelling: it is distribution free, it is intrinsically robust to population outliers and it is very simple to implement. In this section we therefore first describe this approach, before contrasting it with outlier robust mixed modelling in subsequent sections.

To start, we introduce the concept of regression M-quantiles as a general way of indexing a conditional distribution. By definition, regression models are models for a conditional mean $E(y|\boldsymbol{x})$, with mixed models extending this to $E(y_{ij}|\boldsymbol{x}_{ij} = \boldsymbol{x}, j \in i)$. But we could also model $Median(y|\boldsymbol{x})$, or more generally, for $0 < q < 1$, regression quantiles $Q(\boldsymbol{x}_{ij}; q)$, where $Pr(y \leqslant Q(\boldsymbol{x}; q)|\boldsymbol{x}_{ij}) = q$ (Koenker and Bassett Jr, 1978). Newey and Powell (1987) developed a similar generalisation for conditional expectations, introducing regression expectiles $E(\boldsymbol{x}_{ij}; q)$, where $E(\boldsymbol{x}_{ij}; 0.5) = E(y_{ij}|\boldsymbol{x}_{ij})$. Both regression quantiles and regression expectiles can be used to index the support of the conditional distribution of y given $\boldsymbol{x}$ via the quantile index q, and both are special cases of regression M-quantiles $M(\boldsymbol{x}_{ij}; q)$ (Breckling and Chambers, 1988).

Let y denote a continuously distributed scalar random variable with conditional distribution function $F(t|\boldsymbol{x})$. The regression M-quantile of order q for y at $\boldsymbol{x}$ is the value $M(\boldsymbol{x}, q)$ that satisfies

$$\int \boldsymbol{\Psi}_q(\boldsymbol{\sigma}_q^{-1}(t - M(\boldsymbol{x}; q)))d\boldsymbol{F}(t|\boldsymbol{x}) = 0,$$

where $\boldsymbol{\sigma}_q$ is the scale of the random variable $y - M(\boldsymbol{x}; q)$; $\boldsymbol{\Psi}_q(u) = 2\boldsymbol{\Psi}(u)\{qI(u > 0) + (1-q)I(u \leqslant 0)\}$ and $\boldsymbol{\Psi}$ is a user-defined influence function. That is, $M(\boldsymbol{x}; q)$ is a robust M-functional when $\boldsymbol{\Psi}$ is bounded. In this paper we use the Huber-type influence function $\boldsymbol{\Psi}(t) = tI(-c < t < c) + csgn(t)I(|t| \geqslant c)$, so our notation does not distinguish between different types of influence functions. We also note that the value of the tuning constant c will not be explicitly denoted unless we wish to compare different Huber-type influence functions. Thus, when c tends to ∞, so $\boldsymbol{\Psi}(t) = t$, $M(\boldsymbol{x}; q)$ is the regression expectile of order q, and when c tends to 0, so $\boldsymbol{\Psi}(t) = sgn(t)$, $M(\boldsymbol{x}; q)$ is the regression quantile of order q. A linear M-quantile regression model is one where we write

$$M(\boldsymbol{x}_{ij}; q_i) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_q, \tag{11}$$

noting that the parameter $\boldsymbol{\beta}_q$ here, besides depending on q, also depends on $\boldsymbol{\Psi}$, and hence on c.

Given a sample s of n values of $\boldsymbol{x}_{ij}$ and a corresponding set of n independent draws of y from $F(t|\boldsymbol{x})$, an estimate $\hat{\boldsymbol{\beta}}_q$ of $\boldsymbol{\beta}_q$ can be calculated for any value of q in the interval $(0, 1)$ as the solution to the estimating equations

$$n^{-1} \sum_{i \in D} \sum_{j \in s_i} \boldsymbol{\Psi}_q \big(\boldsymbol{s}_q^{-1} (y_{ij} - \boldsymbol{x}_{ij}^T \hat{\boldsymbol{\beta}}_q) \big) \boldsymbol{x}_{ij} = \boldsymbol{0}, \tag{12}$$

where s_q is a robust estimate of σ_q , e.g. the median of the absolute values of the residuals $y_{ij} - \boldsymbol{x}_{ij}^T \hat{\boldsymbol{\beta}}_q$. Provided that the tuning constant c is bounded away from 0, we can easily solve (12) via standard iteratively reweighted least squares (IRLS). The R function *QRLM* that we use to fit regression M-quantiles in this paper is based on the IRLS algorithm underpinning the *rlm* function of the *mass* package in R. It is fast and stable and is available from the authors upon request. An issue that occasionally rises is where two distinct values of q lead to solutions to (12) that 'cross' within the support of the sample $\boldsymbol{x}$ values. This type of behaviour is usually an indication that a linear specification for $M(\boldsymbol{x}; q)$ may not be appropriate. In any case, if one wishes to persist with the linear specification, then the approach described by He (1997) and Salvati et al. (2012) can be used when solving (12).

M-quantile regression models are ensemble models, i.e. they model a range of potential realisations for y given $\boldsymbol{x}$ by fitting (11) for a range of values of

q. To illustrate, in Figure 3 we show the linear regression M-quantile fits for TAXTOT given TURNREG in the ABI data set when $c = 1.345$ (the standard default tuning constant value for a Huber-type influence function) and with q varying from 0.005 to 0.995. By construction, these fits are robust to large residuals since they are based on M-estimates. Figure 3 shows also the line of the least squares fit. This fit is strongly affected by outlying observations with relatively high TAXTOT values. Note also that by fitting linear regression M-quantiles at different values of q (i.e. by fitting an ensemble model) we have a very straightforward way of modelling the heteroskedasticity in the ABI data. Finally, we see that the entire M-quantile regression modelling exercise is distribution free. Our only strong assumption is one of linearity for the different regression M-quantiles. And, though we do not pursue this issue here, even this assumption is not necessary since we can easily modify the M-quantile fits to be non-parametric (Pratesi et al., 2009).

3.2 M-quantile Models for SAE

M-quantile regression ensemble models were suggested as an alternative to mixed models for SAE by Chambers and Tzavidis (2006). Their use in SAE is based on the observation that these ensemble models can be used to create an indexing of the areas not dissimilar to that defined by the random area effects used in mixed models. In particular, if there is significant between area variation in y not explained by between-area variation in $\boldsymbol{x}$, then observations y_{ij} from area i with the same value of $\boldsymbol{x}$ will tend to cluster together in the conditional density $f(y|\boldsymbol{x})$, or equivalently, if we define q_{ij} as the solution to $y_{ij} = M(\boldsymbol{x}_{ij}; q_{ij})$, then the random unit-specific indexes q_{ij} for area i will be correlated. An area-specific index q_i can then be defined as the expected value of these unit-specific indices, and we can model the area i specific regression of y on $\boldsymbol{x}$ by the regression M-quantile $M(\boldsymbol{x}_{ij}; q_i)$. In practice, we substitute estimates for these indices, first by solving $y_{ij} = \widehat{M}(\boldsymbol{x}_{ij}; \hat{q}_{ij})$ for $\hat{q}_{ij}$, where $\widehat{M}(\boldsymbol{x}_{ij}; q) = \boldsymbol{x}^T \hat{\boldsymbol{\beta}}_q$, and then setting $\hat{q}_i = n_i^{-1} \sum_{j \in s_i} \hat{q}_{ij}$. Under this approach,

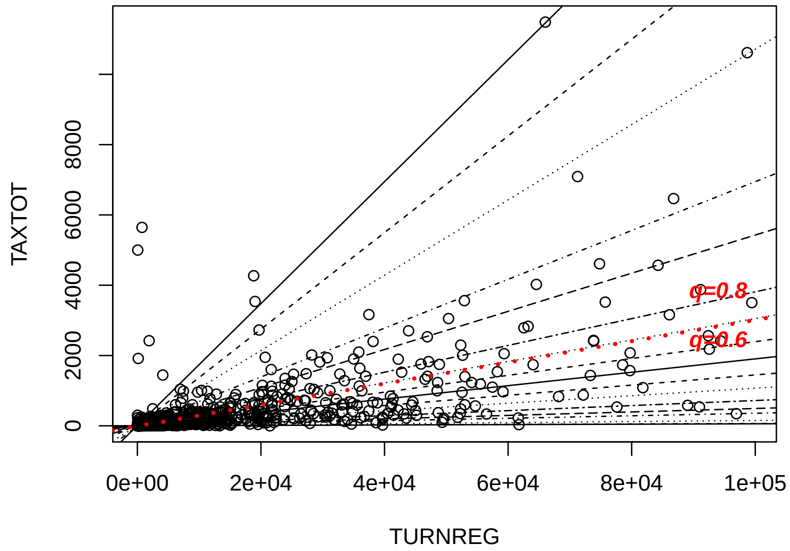


Fig. 3 Regression M-quantile lines for TAXTOT given TURNREG in ABI data with q varying from 0.005 to 0.995. Red bold dotted line is least squares fit. Labels for identifying the q-values close to the OLS-line are reported.

the predicted value of non-sampled unit j in area i is $\widehat{M}(\boldsymbol{x}_{ij}; \hat{q}_i)$, and the corresponding predictor of small area mean $\bar{y}_i$ is

$$MQ_i = N_i^{-1}\Big\{\sum_{j\in s_i} y_{ij} + \sum_{j\in U_i\setminus s_i} \widehat{M}(\boldsymbol{x}_{ij}; \hat{q}_i)\Big\}. \qquad (13)$$

Exactly the same argument as that used to derive the predictor SRB_i (9) as the mean functional defined by the area i distribution function predictor (8) can now be applied in reverse to show that (13) is the mean functional defined by the M-quantile based area i distribution function predictor

$$\hat{F}_i^{MQ}(t) = N_i^{-1}\Big[\sum_{j\in s_i} I(y_{ij} \leqslant t) + \sum_{j\in U_i\setminus s_i} \hat{P}_{ij}^{MQ}(t)\Big], \qquad (14)$$

where $\hat{P}_{ij}^{MQ}(t) = n_i^{-1}\sum_{j\in s_i} I(\widehat{M}(\boldsymbol{x}_{ij}; \hat{q}_i) \leqslant t)$. An MSE estimator for (13), based on the linearization approach of Booth and Hobert (1998), is described in Chambers et al. (2014). This is

$$\widehat{mse}(MQ_i) = \widehat{V}(MQ_i) + \left\{\widehat{B}(MQ_i)\right\}^2, \tag{15}$$

where

- $\widehat{V}(MQ_i) = (1 - n_i N_i^{-1})^2 \left\{\overline{\boldsymbol{x}}_{ri}^T \widehat{V}(\widehat{\boldsymbol{\beta}}_{qi})\overline{\boldsymbol{x}}_{ri}\right\} + (1 - n_i N_i^{-1})^2 \widehat{V}(\overline{e}_{ri})$;
- $\widehat{B}(MQ_i) = N_i^{-1}\left\{\sum_k \sum_{j\in s_k} w_{kj}\boldsymbol{x}_{kj}^T \widehat{\boldsymbol{\beta}}_{\widehat{q}_k} - \sum_{j\in s_i} \boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}}_{\widehat{q}_i}\right\}$;
- $\widehat{V}(\overline{e}_{ri}) = (N_i - n_i)^{-1}(n-1)^{-1}\sum_k \sum_{j\in s_k}\left(y_{kj} - \boldsymbol{x}_{kj}^T\widehat{\boldsymbol{\beta}}_{\widehat{q}_k}\right)^2$.

Here $\widehat{V}(\widehat{\boldsymbol{\beta}}_{qi})$ is the estimated variance of the fitted M-quantile regression coefficients at $q = \widehat{q}_i$, $w_{ij} = b_{ij} + N_i n_i^{-1} I(j \in i)$, $\boldsymbol{b}_i = (b_{ij}) = \boldsymbol{W}(\widehat{q}_i)\boldsymbol{X}_s(\boldsymbol{X}_s^T\boldsymbol{W}(\widehat{q}_i)\boldsymbol{X}_s)^{-1}(N_i - n_i)(\overline{\boldsymbol{x}}_{ri} - \overline{\boldsymbol{x}}_{si})$ and $\boldsymbol{W}(\widehat{q}_i)$ is the diagonal weight matrix obtained from IRLS procedure to calculate $\widehat{\boldsymbol{\beta}}_{\widehat{q}_i}$.

As with the SR_i predictor, the predictor MQ_i defined by (13) is robust projective since it assumes all non-sampled units follow the working model defined by the M-quantile indexing of the different areas of interest. Given individual $\boldsymbol{x}$ values, this model characterises the y values of area i units as satisfying

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{q_{ij}} = \boldsymbol{x}_{ij}^T(\boldsymbol{\beta}_{q_i} + \boldsymbol{\gamma}_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{q_i} + u_{ij}, \tag{16}$$

where the $\boldsymbol{\gamma}_{ij}$ are noise variables with zero expectation over area i, but with unspecified variability. As a consequence, the errors u_{ij} are heteroskedastic. Tzavidis et al. (2010) note that this heteroskedasticity implies that (14) is then typically biased, and consequently so is (13). They propose replacing (13) by the mean functional of a bias-corrected M-quantile based predictor for the area i distribution function for y, extending the approach of Welsh and Ronchetti (1998) to robust prediction based on M-quantile modelling. This replaces $\widehat{P}_{ij}^{MQ}(t)$ in (14) by a robust bias-corrected alternative

$$\widehat{P}_{ij}^{MQB}(t) = n_i^{-1}\sum_{k\in s_i} I\left(\widehat{M}(\boldsymbol{x}_{ij};\widehat{\boldsymbol{q}}_i) + \widehat{\boldsymbol{\omega}}_i^{MQ}\boldsymbol{\Phi}\{\boldsymbol{r}_{ik}^{MQ}/\widehat{\boldsymbol{\omega}}_i^{MQ}\} \leqslant t\right),$$

where $\boldsymbol{r}_{ik}^{MQ} = y_{ik} - \widehat{M}(\boldsymbol{x}_{ik};\widehat{q}_i)$, $\boldsymbol{\Phi}$ is another bounded influence function that satisfies $|\boldsymbol{\Phi}| \geqslant |\boldsymbol{\Psi}|$ and $\widehat{\boldsymbol{\omega}}_i^{MQ}$ is a robust estimator of the scale of the residuals

r_{ik}^{MQ} in area i. The mean functional defined by this alternative distribution function estimator is then a corresponding bias-corrected version of (13) of the form

$$\begin{aligned} MQB_i &= N_i^{-1}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in U_i\setminus s_i} \widehat{M}(\boldsymbol{x}_{ij};\hat{q}_i) + n_i^{-1}(N_i - n_i)\sum_{j\in s_i} \boldsymbol{\omega}_{ij}^{MQ}\boldsymbol{\Phi}\Big\{\big(y_{ij} - \widehat{M}(\boldsymbol{x}_{ij};\hat{q}_i)\big)\Big/\boldsymbol{\omega}_{ij}^{MQ}\Big\}\right] \\ &= MQ_i + (1 - n_iN_i^{-1})n_i^{-1}\sum_{j\in s_i} \boldsymbol{\omega}_{ij}^{MQ}\boldsymbol{\Phi}\Big\{\boldsymbol{r}_{ij}^{MQ}/\boldsymbol{\omega}_{ij}^{MQ}\Big\}. \end{aligned} \tag{17}$$

Here $\boldsymbol{\omega}_{ij}^{MQ}$ is a robust estimator of the scale of the residual r_{ij}^{MQ} in area i. The use of the two influence functions, $\boldsymbol{\Psi}$ used to define the working M-quantile model (16), see (12), and $\boldsymbol{\Phi}$ in the bias correction term in (17), is worthy of comment. The aim of the $\boldsymbol{\Psi}$ function is to ensure that sample outliers have little or no influence on the fit of the working M-quantile model. The second, $\boldsymbol{\Phi}$, is still bounded but 'less restrictive' than $\boldsymbol{\Psi}$ (since $|\boldsymbol{\Phi}| \geqslant |\boldsymbol{\Psi}|$), and its purpose is to define an adjustment for the bias caused by the fact that the first two terms on the right hand side of (17) treat sample outliers as not being representative. Finally, we note that the MSE of (17) can also be estimated via CST (Chambers et al., 2014), which in this case is based on the approximation

$$\begin{aligned} mse(MQB_i) =&(1 - n_iN_i^{-1})^2\times \\ &\left[\Big\{\overline{\boldsymbol{x}}_{ri} - \overline{\boldsymbol{x}}_{si}\Big\}^T\hat{\boldsymbol{V}}\big(\hat{\boldsymbol{\beta}}_{\hat{q}_i}\big)\big\{\overline{\boldsymbol{x}}_{ri} - \overline{\boldsymbol{x}}_{si}\big\} + \hat{\boldsymbol{V}}(\overline{e}_{ri}) + \frac{1}{n_i^2}\sum_{j\in s_i}\Big\{\boldsymbol{\omega}_{ij}^{MQ}\boldsymbol{\Phi}\big((\boldsymbol{\omega}_{ij}^{MQ})^{-1}r_{ij}^{MQ}\big)\Big\}^2\right]. \end{aligned} \tag{18}$$

Note that in expression (18) there is no squared bias term, since this bias is (approximately) removed by the bias correction term of MBQ. An alternative pseudo-linearization estimator of the conditional MSE of MQ and MQB has been developed in Chambers et al. (2014), applying the ideas that are set out in Chambers et al. (2011).

4 Outlier robust bootstrapping via the Bounded Block Bootstrap in SAE

The bootstrap technique, see Efron (1992) and Tibshirani and Efron (1993), is a computational approach to sample inference that provides answers to a large class of statistical problems. In particular, its main use is in confidence interval estimation. It was originally developed for parametric inference given i.i.d. data and it often recommended as an alternative when distributional assumptions are questionable (Davison et al., 1997). However, random effects models for hierarchically dependent data, e.g. clustered or multilevel data, are now in wide use. With such data, it is important to use bootstrap techniques that replicate the hierarchical dependence structure of the data under the mixed model, typically via a percentile bootstrap with resampling of scaled level one (unit level) and level two (area level) residuals (Fama and French, 1988).

A standard approach to SAE for means or totals uses a working model with mixed effects like (1) to characterise between area variability. Bootstrap estimation of the MSE and of confidence intervals for a predictor like (3) is straightforward when it is based on this working model, i.e. when resampling is carried out from simulated population data based on this model. However, the validity of this parametric bootstrap is dependent on the working model providing an adequate fit to the unobserved population values. In particular, this bootstrap cannot replicate the true population structure if outliers in the sample data imply that the stochastic assumptions of the working model (1) are questionable. Substitution of outlier robust parameter estimates leads to biased estimates of variance components that can in turn lead to a biased parametric bootstrap. Use of a non-parametric bootstrap also has issues in this case since with replacement resampling from working model-based empirical residuals that include outlier values can lead to bootstrap error distributions that are too wide.

In what follows, we propose an M-quantile ensemble modelling approach to recreating the hierarchical population variability in the original population, combining it with a modified block bootstrap that allows outliers to have a controlled effect on the bootstrap variability. Chambers and Chandra (2013) introduced a random effects block bootstrap that allows arbitrary level one error distributions, but assumes exchangeability at level two. We use an adaptation of this bootstrap, which we refer to as a bounded block bootstrap (BBB). This is a non-parametric bootstrap procedure that replicates hierarchical data structure, but uses 'Huberized' level one and level two working model residuals to restrict the influence of outliers on the bootstrap populations and a block bootstrap resampling procedure to ensure that actual level one (unit level) variability within areas is replicated. The steps in this bootstrap procedure are as follows:

1. Calculate the robust estimate $\hat{\boldsymbol{\beta}}_{0.5}$ for the fixed effects vector defining the linear regression M-quantile of order $q = 0.5$, as well as the marginal residuals defined by this fit $r_{ij} = y_{ij} - \boldsymbol{x}_{ij}^T \hat{\boldsymbol{\beta}}_{0.5}$;
2. Calculate the level two and level one empirical residuals generated by the fitted linear quantile regression of order $q = 0.5$ as $r_i^{(2)} = n_i^{-1} \sum_{j \in s_i} r_{ij}$ and $r_{ij}^{(1)} = r_{ij} - r_i^{(2)}$ respectively;
3. Calculate the moment-based estimates $\tilde{\sigma}_u^2$ and $\tilde{\sigma}_e^2$ of the between area and within area variance components defined by the hierarchical linear model (1), e.g. using the *mbest* package within R, and then scale the level two and level one empirical residuals calculated in the previous step so that their sample variances are equal to $\tilde{\sigma}_u^2$ and $\tilde{\sigma}_e^2$ respectively;
4. Compute 'Huberized' versions of these level two and level one empirical residuals

$$\breve{r}_i^{(2)} = r_i^{(2)} I\big(|r_i^{(2)}| \leqslant c\tilde{\sigma}_u^2\big) + c\tilde{\sigma}_u^2 sgn\big(r_i^{(2)}\big) I\big(|r_i^{(2)}| > c\tilde{\sigma}_u^2\big)$$
$$\breve{r}_{ij}^{(1)} = r_{ij}^{(1)} I\big(|r_{ij}^{(1)}| \leqslant c\tilde{\sigma}_e^2\big) + c\tilde{\sigma}_e^2 sgn\big(r_{ij}^{(1)}\big) I\big(|r_{ij}^{(1)}| > c\tilde{\sigma}_e^2\big),$$

and then mean correct these bounded residuals. The tuning constant c here is set high enough to allow outliers to have an impact on the residuals. We set $c = 3$ for the results reported in this paper, but larger values (e.g. $c = 6$) could also be used;

5. Implement a two level random effects block bootstrap based on these bounded and mean corrected empirical residuals. That is, for each bootstrap iteration b,

   a. Generate level two errors for the D areas by drawing a simple random sample of size D with replacement from the set $\{\breve{r}_i^{(2)}; i = 1, \ldots, D\}$ of bounded and mean corrected empirical level two residuals. Denote this sample by $\{\breve{r}_i^{(*2)}; i = 1, \ldots, D\}$. Similarly, generate level one errors within each area i by independently drawing simple random samples of size N_i with replacement from $\{\breve{r}^1_{i*j}; j \in U_{i*}\}$ where i^* is a random drawn from $1, \ldots, D$. Denote this sample by $\{\breve{r}_{ij}^{(*1)}; i = 1, \ldots, N_i\}$;

   b. Simulate bootstrap population data $(y^*_{ij}, \boldsymbol{x}_{ij})$ using $y^*_{ij} = \boldsymbol{x}_{ij}\widehat{\boldsymbol{\beta}}_{0.5} + \breve{r}_i^{(*2)} + \breve{r}_{ij}^{(*1)}$ and compute the corresponding bootstrap small area mean $\overline{y}_i^{(b)}$ at iteration b;

   c. draw a bootstrap sample from the bootstrap population using the sampling method used to obtain the original sample;

   d. Fit a linear M-quantile regression ensemble model to the bootstrap sample data drawn in step 3 to obtain bootstrap estimates for the model parameters and for the individual M-quantile indices $\widehat{q}_{ij}$ and the area-level indices $\widehat{q}_i$. Use these to compute the bootstrap M-quantile estimates $MQ_i^{(b)}$ and $MQB_i^{(b)}$ for $\overline{y}_i^{(b)}$ at bootstrap iteration b;

   e. Repeat steps a-d B times to obtain B sets of bootstrap small area estimates. The MSE estimates for MQ_i and MQB_i are then computed as

$$mse_i(MQ_i) = \frac{1}{B}\sum_{b=1}^{B}\left(MQ_i^{(b)} - \overline{y}_i^{(b)}\right)^2 \tag{19}$$

   and

$$mse_i(MQB_i) = \frac{1}{B}\sum_{b=1}^{B}\left(MQB_i^{(b)} - \overline{y}_i^{(b)}\right)^2. \tag{20}$$

   Confidence intervals can also be computed from this bootstrap distribution, again using the percentile method.

5 Analysis of the ABI data

5.1 Initial Raw Scale Modelling

In order to apply the outlier robust methods discussed in Sections 2 - 4, sample data for TAXTOT and TURNREG as well as register values of TURNREG are required. These register values are not available for the ABI sample data. Consequently, the sample values of TURNREG were independently bootstrapped within each level of CLASS in order to create a synthetic register containing the same number of businesses as the estimated CLASS populations shown in Table 1. This synthetic register was then held fixed for the purpose of illustration, with the target being estimation of the CLASS averages of TAXTOT.

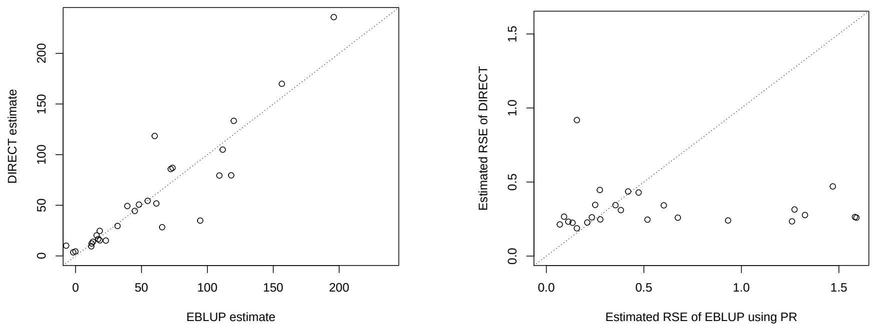
To start, values of the direct estimator (here denoted DIRECT) as well as that of the EBLUP (3) were calculated for each CLASS group. Since the synthetic register is, by construction, almost perfectly balanced on TURNREG within each CLASS group, DIRECT was defined as the CLASS level sample mean of TAXTOT. Figure 4 illustrates the relationship between the CLASS values for DIRECT and EBLUP by showing the scatterplots of the direct estimates and the EBLUP estimates and their relative standard errors (RSEs). The relative standard error (Gonzalez et al., 1975) is defined here as:

$$RSE = \frac{\sqrt{\widehat{mse}(\hat{\theta})}}{\hat{\theta}} \times 100,$$

where $\hat{\theta}$ is an estimator of actual mean θ. We see that values of EBLUP and DIRECT generally line up in the left hand plot in Figure 4, but there are some notable differences. We also note that there is no obvious shrinkage, indicating that there may be a problem with the EBLUP estimates. Inspection of the ABI values for TAXTOT within different CLASS groups shows two major issues. The first is that the EBLUP estimates for CLASS = 31, 71 and 72 are negative, with extremely high estimated RSEs. The second was noted earlier in Section 1, and corresponds to the presence of a large number of outlier values in the ABI. Their impact on the corresponding EBLUP estimates can be seen in the right hand plot in Figure 4, which shows that the estimated EBLUP RSEs for over half of the 28 CLASS groups are greater than the corresponding RSEs of the DIRECT estimator. We have also computed the MSE estimates of the EBLUP using a parametric bootstrap procedure proposed by González-Manteiga et al. (2007). The empirical correlation between the RSEs calculated via the PR and the bootstrap methods is 0.99. The two procedures give practically identical results, but the bootstrap procedure has lower computational time than the PR method (0.22 vs. 24.71 minutes using optimised R code on a laptop with a 2 GHz Intel Core i7 and 16 Gb RAM).

We also calculated the SR predictors using the package rsae (Schoch, 2014). Here we noted that the algorithm used to calculate these estimate often failed to converge, and, even when it did converge, the estimates that were produced were substantially smaller than the direct estimates (Figure 5). This could be due to the fact that the predicted area effects were almost zero under this approach.

We therefore focussed on the use of an M-quantile modeling approach for estimation of the CLASS means of TAXTOT. The fit of this model to the ABI data is shown in Figure 6, while Figure 7 shows how the CLASS level estimates generated by MQ and MQB based on this fitted M-quantile model compare with the corresponding DIRECT estimates. More precisely, Figure 6 shows the M-quantile regression fits to the different CLASS groups. These fits show how the M-quantile regression model captures the between area variation in the response variable (TAXTOT) that not explained by the covariate (TURNREG). We can also note that the slope of the fixed effects fitted line based on the linear mixed model (dashed line in Figure 6) is higher than the corresponding slopes of the different CLASS level M-quantile regression fits. This suggests that outlying data values for TAXTOT in the ABI data lead to



(a) (b)

Fig. 4 Left plot (a) shows DIRECT vs. EBLUP estimates for ABI sample. Dotted line is the identity line. Right plot (b) shows corresponding estimated relative standard errors for DIRECT vs. those for EBLUP, calculated via the PR method, with estimated relative standard errors for three CLASS groups with negative EBLUP estimates excluded.

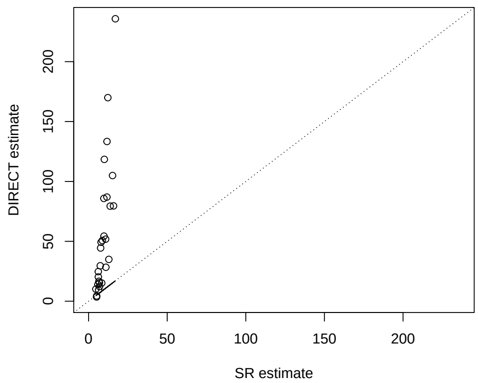


Fig. 5 DIRECT vs SR estimates for ABI sample. Dotted line is the identity line.

biased CLASS estimates when a linear mixed model is fitted to these data.

Figure 7 displays bias diagnostic plots (left column) where CLASS values of DIRECT have been plotted against corresponding values of MQ and MQB. Athough there is evidence of shrinkage, the MQ and MQB estimates appear

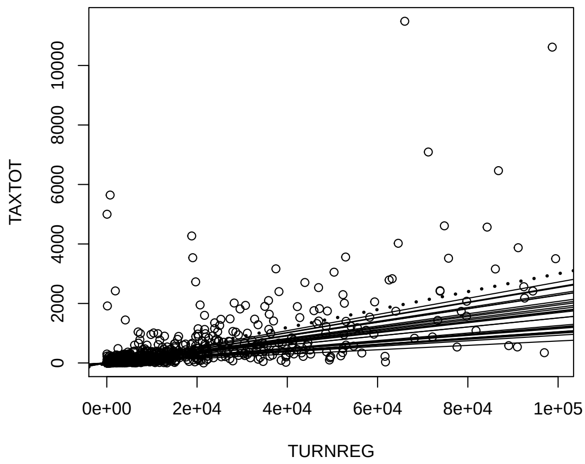


Fig. 6 Linear regression M-quantile fits to the CLASS groups. Dashed line is fixed effects fitted values from linear mixed model underpinning the EBLUP estimates.

to be generally consistent with the direct estimates. However, there are five CLASS values for which the MQ, MQB and direct estimates are notably different. The right column of Figure 7 shows the estimated RSEs for DIRECT plotted against the corresponding estimated RSEs for MQ and MQB, where the latter are based on the analytic MSE estimator CST. As already noted, there are large number of outlier values in the ABI, but their impact on the corresponding MQ and MQB estimates is smaller than their impact on the EBLUP estimates: the estimated RSEs of DIRECT are greater than the corresponding estimated RSEs of the MQ and MQB estimators in over half of the 28 CLASS groups, with the estimated RSEs of DIRECT lower than the estimated RSEs of MQB in just 7 out of the 28 CLASS groups. This provides some evidence that the bias correction used in MQB improves the performance of the MQ predictor.

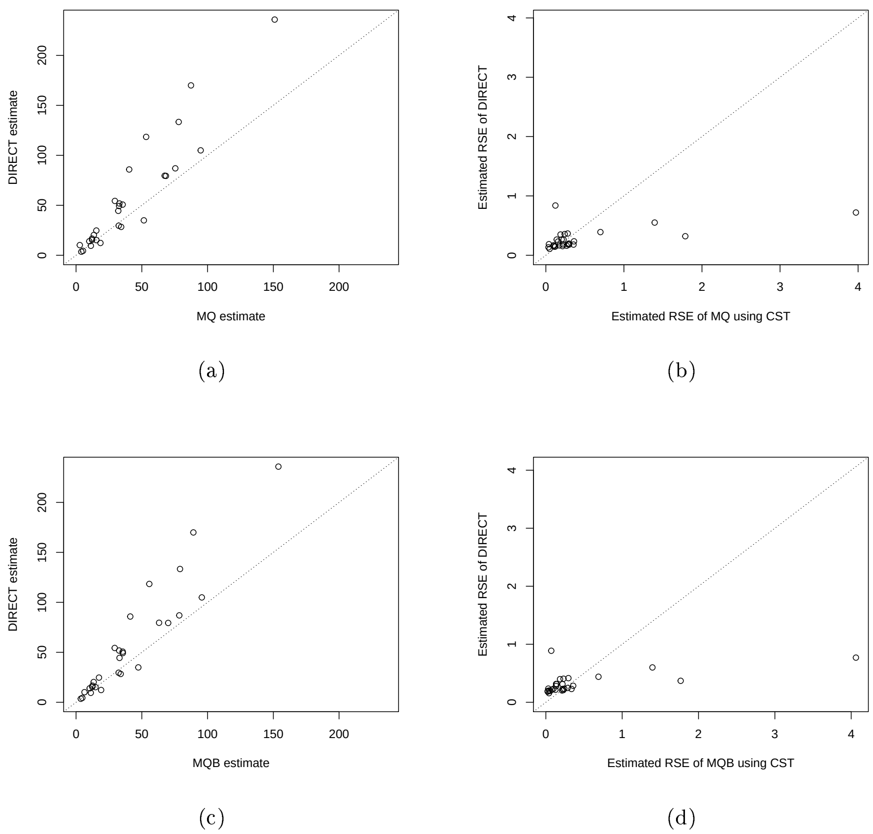


Fig. 7 CLASS values of DIRECT plotted against corresponding values of MQ and MQB (left column) and estimated RSEs for DIRECT plotted against corresponding RSEs for MQ and MQB (right column). Note that the dotted line is the identity line. Estimated RSEs for MQ and MQB are based on the analytic MSE estimator CST.

5.2 A Log Scale Bounded Block Bootstrap

A log-log linear specification is a standard way of modelling income and expenditure data in economics (Emerson and Stoto, 1983). This is because most economic variables exhibit multiplicative, rather than additive, behaviour (Oltean, 2016). As a consequence, we now fit an M-quantile regression ensemble model to log-scale ABI data. The log-scale-M-quantile regression fits to the CLASS groups are shown in Figure 8. There still appear to be some outlying

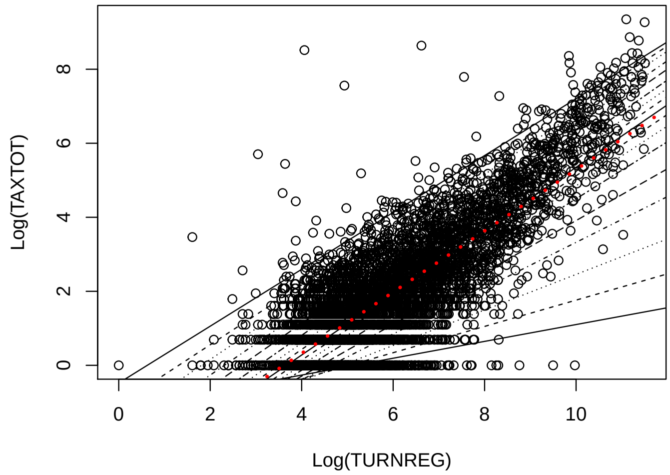
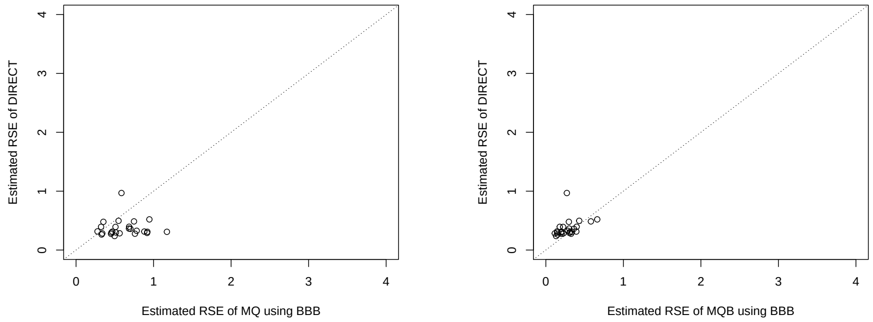


Fig. 8 Log scale regression M-quantile lines (q = 0.005, 0.01, 0.02,0. 05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.98, 0.99, 0.995) for TAXTOT given TURNREG in ABI data. Red bold dotted line is least squares fit.

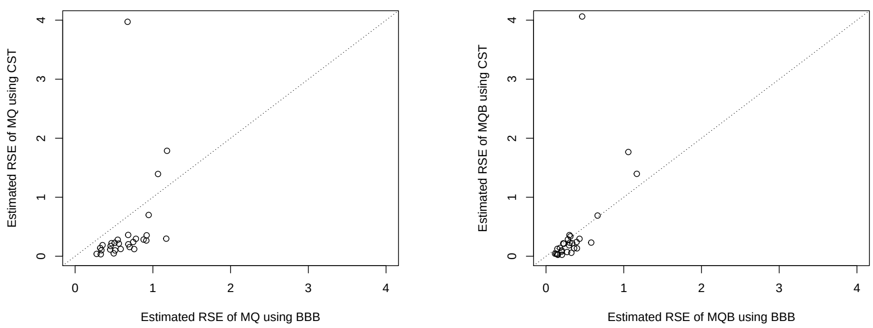
data points, however, and it is possible that these values can lead to biased small area estimates using a log-scale mixed model. Consequently, we again consider outlier robust small area estimators based on M-quantile regression, but now applied to log-scale ABI data.

Figure 9 shows the estimated RSEs for DIRECT plotted against corresponding RSEs for MQ and MQB where the estimated RSEs for MQ and MQB are based on a log-scale Bounded Block Bootstrap (see Section 4), followed by exponentiation of the log-scale bootstrapped population values. There are a large number of outlier values in the ABI, but their impact on the bootstrap MSEs for MQ and MQB is smaller than their impact on corresponding MSE estimates derived using the analytic MSE estimator CST (Figure 10).



(a) (b)

Fig. 9 Estimated RSEs of MQ (a) and MQB (b) computed via log scale bounded block bootstrap compared with estimated RSEs generated by DIRECT for CLASS groups in ABI data.



(a) (b)

Fig. 10 Estimated RSEs of MQ (a) and MQB (b) computed via log scale bounded block bootstrap compared with estimated RSEs of MQ and MQB based on the analytic MSE estimator CST for CLASS groups in ABI data.

6 Concluding remarks

In this paper we have examined how the presence of outliers in business and income surveys can impact on small area estimates calculated using data from these surveys. After a review of recent developments in outlier robust small domain estimation based on the projective and predictive approaches, we have proposed an outlier robust bootstrap MSE estimator for small area predictors based on the M-quantile modelling approach. We focused on this approach because M-quantile modelling is fast and stable, and offers a practical approach to robust SAE for business surveys. We then illustrated this approach using data that closely mimics data that were observed in the UK ABI.

The empirical results that are reported in Section 5, and in the model-based and design-based simulations reported in the Supplementary Materials, show that the bias corrected version of the M-quantile method works quite well. In particular, when there are significant departures from the assumed working small area model, the robust-predictive estimators defined via this approach are less biased and can be more efficient than the robust-projective estimators associated with the standard M-quantile approach.

The CST linearisation MSE estimator offers a promising approach to analytic estimation of the MSE of robust-predictive estimators: it shows small bias but has problems with under-coverage (see the results in Supplementary Materials for model and design based simulations). This could occur because the CST estimator is based on an unconditional approach, and its variance and squared bias terms suffer from instability at small sample sizes. It is also reflected in the unstable coverage behaviour of the nominal 95% confidence intervals that are generated using CST. To overcome this problem a Bounded Block Bootstrap approach to estimating the MSEs of MQ-based SAE predictors has been proposed. It is very stable under all simulation scenarios described in the Supplementary Materials, but also displays positive bias with increasing heteroskedasticity due to outlying values in both area and individual effects. In this context we conjecture that a larger value of the tuning constant used to compute the 'Huberized' versions of the level two and level one empirical residuals used in the block bootstrap procedure could potentially reduce this bias. However, this requires further research. We also note that the proposed bootstrap procedure improves on the coverage rate of nominal 95% confidence intervals generated by the CST MSE estimator.

Finally, we note that the good performance of the Bounded Block Bootstrap procedure for MQ-based SAE predictors does suggest extension of this method to the SR and SRB predictors. This is left for further research, together with the development of algorithms for fitting robust mixed models that exhibit improved convergence behaviour (Schoch, 2012).

References

Battese GE, Harter RM, Fuller WA (1988) An error-components model for prediction of county crop areas using survey and satellite data. Journal of the American Statistical Association 83(401):28–36

Booth JG, Hobert JP (1998) Standard errors of prediction in generalized linear mixed models. Journal of the American Statistical Association 93(441):262–272

Breckling J, Chambers R (1988) M-quantiles. Biometrika 75(4):761–771

Chambers R, Chandra H (2013) A random effect block bootstrap for clustered data. Journal of Computational and Graphical Statistics 22(2):452–470

Chambers R, Tzavidis N (2006) M-quantile models for small area estimation. Biometrika 93(2):255–268

Chambers R, Chandra H, Tzavidis N (2011) On bias-robust mean squared error estimation for pseudo-linear small area estimators. Survey Methodology 37:153–170

Chambers R, Chandra H, Salvati N, Tzavidis N (2014) Outlier robust small area estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76(1):47–69

Chambers RL (1986) Outlier robust finite population estimation. Journal of the American Statistical Association 81(396):1063–1069

Davison AC, Hinkley DV, et al. (1997) Bootstrap methods and their application, vol 1. Cambridge university press

Efron B (1992) Bootstrap methods: another look at the jackknife. In: Breakthroughs in statistics, Springer, pp 569–593

Emerson J, Stoto M (1983) Understanding robust and exploratory data analysis. New York: John Wiley

Fama EF, French KR (1988) Dividend yields and expected stock returns. Journal of financial economics 22(1):3–25

Fellner WH (1986) Robust estimation of variance components. Technometrics 28(1):51–60

Giusti C, Pratesi M, Stefano M, Salvati N (2012) Robust small area estimation and oversampling in the estimation of poverty indicators. Survey Research Methods 6(3):155–163

Gonzalez ME, Ogus JL, Shapiro G, Tepping BJ (1975) Standards for discussion and presentation of errors in survey and census data. Journal of the American Statistical Association 70(351b):5–23

González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2007) Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. Computational statistics & data analysis 51(5):2720–2733

Hall P, Maiti T (2006) On parametric bootstrap methods for small area prediction. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68(2):221–238

He X (1997) Quantile curves without crossing. The American Statistician 51(2):186–192

Jiongo VD, Haziza D, Duchesne P (2013) Controlling the bias of robust small-area estimators. Biometrika 100(4):843–858

Koenker R, Bassett Jr G (1978) Regression quantiles. Econometrica: journal of the Econometric Society pp 33–50

Newey WK, Powell JL (1987) Asymmetric least squares estimation and testing. Econometrica: Journal of the Econometric Society pp 819–847

Oltean E (2016) Modelling income, wealth, and expenditure data by use of econophysics - phd thesis. arXiv preprint arXiv:160308383

Prasad NN, Rao JN (1990) The estimation of the mean squared error of small-area estimators. Journal of the American statistical association 85(409):163–171

Pratesi M, Ranalli MG, Salvati N (2009) Nonparametric m-quantile regression using penalised splines. Journal of Nonparametric Statistics 21(3):287–304

Rao J, Molina I (2015) Small Area Estimation. Wiley, New York

Salvati N, Tzavidis N, Pratesi M, Chambers R (2012) Small area estimation via m-quantile geographically weighted regression. Test 21(1):1–28

Schoch T (2012) Robust unit-level small area estimation: a fast algorithm for large data sets. Austrian Journal of Statistics 41(4):243–265

Schoch T (2014) rsae: Robust Small Area Estimation. R package version 0.1-5

Sinha SK, Rao J (2009) Robust small area estimation. Canadian Journal of Statistics 37(3):381–399

Street JO, Carroll RJ, Ruppert D (1988) A note on computing robust regression estimates via iteratively reweighted least squares. The American Statistician 42(2):152–154

Tibshirani RJ, Efron B (1993) An introduction to the bootstrap. Monographs on statistics and applied probability 57:1–436

Tzavidis N, Marchetti S, Chambers R (2010) Robust estimation of small area means and quantiles. Australian and New Zealand Journal of Statistics 52:167–186

Welsh AH, Ronchetti E (1998) Bias-calibrated estimation from sample surveys containing outliers. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60(2):413–428