Journal of Official Statistics, Vol. 28, No. 1, 2012, pp. 89–106

Constrained Small Area Estimators Based on M-quantile Methods

Enrico Fabrizi[1], *Nicola Salvati*[2], *and Monica Pratesi*[3]

Small area estimators associated with M-quantile regression methods have been recently proposed by Chambers and Tzavidis (2006). These estimators do not rely on normality or other distributional assumptions, do not require explicit modelling of the random components of the model and are robust with respect to outliers and influential observations. In this article we consider two remaining problems which are relevant to practical applications. The first is benchmarking, that is the consistency of a collection of small area estimates with a reliable estimate obtained according to ordinary design-based methods for the union of the areas. The second is the correction of the under/over-shrinkage of small area estimators. In fact, it is often the case that, if we consider a collection of small area estimates, they misrepresent the variability of the underlying "ensemble" of population parameters. We propose benchmarked M-quantile estimators to solve the first problem, while for the second we propose an algorithm that is quite similar to the one used to obtain Constrained Empirical Bayes estimators, but that, consistently with the principles of M-estimation, does not make use of distributional assumptions and tries to achieve robustness with respect to the presence of outliers. The article is essentially about point estimation; we also introduce estimators of the mean squared error, but we do not deal with interval estimation.

Key words: Over-shrinkage; benchmarking; robust estimation.

1. Introduction

In statistical inference about a finite population, estimates of population descriptive quantities for a target variable y are usually needed for the population as a whole and for different collections of subpopulations (domains or areas). A small area estimation problem arises when the available samples are not large enough to allow for reliable estimation using standard design-based methods for all or most of the domains (areas) being considered. A large and growing literature is devoted to this subject; see Rao (2003) and Jiang and Lahiri (2006) for general introduction and recent reviews of the literature.

[1] DISES, Università Cattolica del S. Cuore, Piacenza, Italy. Email: enrico.fabrizi@unicatt.it
[2] Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa, Italy. Email: salvati@ec.unipi.it
[3] Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa, Italy. Email: m.pratesi@ec.unipi.it
Acknowledgments: This article is based on a conference presentation at the ITACOSM, Siena, Italy, 2009. The work of Salvati and Pratesi is supported by the project PRIN 2007 "Efficient use of auxiliary information at the design and at the estimation stage of complex surveys: methodological aspects and applications for producing official statistics" awarded by the Italian Government to the Universities of Perugia, Cassino, Florence, Pisa and Trieste. The work of Salvati and Pratesi is also supported by the project SAMPLE "Small Area Methods for Poverty and Living Condition Estimates" (www.sample-project.eu), financed by the European Commission under the 7th FP.

All small area estimation methods are based on the availability of population-level auxiliary information to improve the precision of the estimation. They have their own specificity in the way of linking auxiliary information and target variable and in the properties of the obtained small area predictors.

When the samples available for each area are very small, model-based predictors are popular. Among these methods, Best and Empirical Best (EB) predictors have become the "standard of the industry" estimators. Let's suppose for simplicity's sake, that a linear model linking the target variable y to a set of auxiliary variables x is plausible. EB predictors are based on the assumption of a linear mixed model in which random effects are introduced to account for the correlation of residuals within the same area. Common criticisms of the EB based on linear mixed models are that they require explicit assumptions on the random effects and that the estimation of the parameters relying on normality or on least squares is sensitive to the presence of outliers or influential observations in the data, a situation that is likely to occur in the analysis of survey data. This limitation can be overcome using Robust EB (Sinha and Rao 2009). Nonetheless, here we prefer to focus on the alternative approach based on M-quantile regression.

Small area estimation based on linear quantile regression and M-estimation has been recently proposed by Chambers and Tzavidis (2006). M-quantile estimation is based on the assumption of a linear relationship between y and x at each quantile of the $y|x$ distribution, but is free of any distributional assumption, is robust with respect to the presence of outliers and influential observations and does not require explicit specification of the random part of the model. How M-quantile regression may be used to obtain small area predictors is reviewed in Section 2.

Model-based methods may not satisfy coherence properties, that may be relevant to final users of small area estimates. In this article we focus on two of these properties. The first is "benchmarking," while the second may be labelled as "neutral shrinkage". Let the small areas be a partition of a larger area. A set of estimates is said to be benchmarking if the estimated totals of y for the small areas sum to the total estimated for the larger area (typically using design unbiased or design consistent methods). The EB predictors do not fulfill the benchmarking property (see Rao 2003, Section 7.2.7 for a discussion of this problem and also of adjusted predictors). As may be expected, this is true also for M-quantile regression based estimators; in fact they are model-based and do not incorporate the sampling weights. We propose a modification of the M-quantile (MQ) predictors estimation algorithm to obtain benchmarked estimates. Formally they will be constrained optimal MQ estimators and will be referred to as "benchmarked MQ" (BMQ).

The second coherence property we consider is "neutral shrinkage", which is a special "ensemble" property, that means a property related to the estimation of a functional of an ensemble of parameters (Frey and Cressie 2003). Specifically, we focus on the estimation of the variance of the underlying population means or totals of y pertaining to an ensemble of small areas, a problem often considered in the literature (see, for instance, Ghosh 1992; Judkins and Liu 2000; Ugarte et al. 2009). An ensemble of estimators has neutral shrinkage if the variance of the ensemble of the parameters can be unbiasedly estimated by the variance of the ensemble of the estimators.

Design-based estimates are typically over-dispersed (they are more spread than the actual population parameters), while EB predictors are under-dispersed, that is they over-shrink.

The behaviour of MQ predictors in this respect has not been studied in the literature. By means of simulations, it may be easily shown that they can either over- or under-shrink depending on the actual distribution of the underlying parameters, as will be made clear in later sections.

In the article we propose an adjustment of the benchmarked MQ predictors in order to obtain estimators with approximately neutral shrinkage. This adjustment parallels the one used to adjust EB predictors (Rao 2003, Section 9.6).

The structure of the article is as follows. In Section 2 MQ estimators are reviewed; in Section 3 we propose benchmarked MQ estimators and benchmarked MQ estimators with approximately neutral shrinkage. A simulation exercise is introduced and its results discussed in Section 4. Section 5 is devoted to the description of the application of the method to a well-known data set, which will test the method also in the presence of outliers. Concluding remarks are contained in Section 6.

2. Small Area Estimators Based on M-quantile Methods

Let us suppose that a population U of size N is divided into m nonoverlapping subsets (domains or areas of interest) of size N_i, $i = 1, \ldots, m$. We are interested in a target variable y and more specifically in estimating the area level means

$$\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$$

Suppose that a random sample is drawn from the population, so that area-specific samples of size $n_i > 0$ are available. It may also be the case that $n_i = 0$ for some areas. The problem of estimation (and benchmarking) for these areas will be addressed at the end of Section 3.1. Values of y are known only for sampled values, but we assume that a vector of p auxiliary variables $\mathbf{x}_{ij}$ is known for each unit in the population. We use subscript of i to denote restriction to small area i, so that s_i (r_i) denotes the set of sample (nonsample) population units from area i, and $U_i = s_i \cup r_i$ denotes the set of population units making up the small area i.

A recently proposed approach to small area estimation is based on the use of M-quantile models (see Chambers and Tzavidis 2006, Section 4). Since much of the development in this article is based on the application of linear quantile/M-quantile regression, we now give a brief definition of these concepts.

Ordinary linear regression is based on the idea of modelling the expected value of the dependent variable as a function of the regressors; that is, in our notation, on the assumption that $E(y_{ij}|\mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \beta$. In quantile regression it is the qth quantile that is assumed to be a linear function of the auxiliary information, i.e.,

$$Q_q(y_{ij}|\mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \beta(q) \quad q \in (0, 1)$$

This means that a distinct (hyper)plane is fitted to the data for each $q \in (0, 1)$ according to quantile-specific regression coefficients $\beta(q)$. See Koenker and Bassett (1978) for a general introduction to quantile regression. The vector of $\beta(q)$ may be estimated according to some minimization criterion such as least absolute deviations considered in Koenker and D'Orey (1987). Breckling and Chambers (1988) introduced the application of robust M-estimation to quantile regression. M-quantile regression provides a "quantile-like" generalization of

regression based on influence functions. For specified q and influence function ψ, an estimate of the vector of the regression parameters $\beta_\psi(q)$ may then be obtained by solving the following normal equations

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\psi_q\Big(y_{ij}-\mathbf{x}_{ij}^T\beta_\psi(q)\Big)\mathbf{x}_{ij}=0$$

in $\beta_\psi(q)$, where $\psi_q(r)=2\psi(s^{-1}r)\{qI(r>0)+(1-q)I(r\le 0)\}$ and $r=(r_{ij})=y_{ij}-\mathbf{x}_{ij}^T\beta_\psi(q)$. The influence function $\psi(\cdot)$ may for instance be chosen to be the Huber proposal 2, i.e., $\psi(u)=uI(-c\le u\le c)+c,\ \mathrm{sgn}\,(u)I(|u|>c),\ u\in\Re$, as we have done in the application in Sections 4 and 5; in this case c is a tuning constant assumed to be bounded away from 0. Consistently with most applications, we set $c=1.345$. This value gives reasonably high efficiency in the normal case; more specifically, it produces 95% efficiency when the errors are normal and still offers protection against outliers (Huber 1981). The quantity s in the definition of $\psi_q(r)$ is a robust estimate of the scale of the data such as the mean absolute deviation $s=med|r|/0.6745$. For specified q an estimate $\hat\beta_\psi(q)$ of $\beta_\psi(q)$ is then obtained via iterative reweighted least squares.

Following Chambers and Tzavidis (2006), an alternative to random effects for characterizing the variability across the population not accounted for by the regressors is to use the M-quantile coefficients of the population units. For unit j in area i, this coefficient is the value θ_{ij} such that $Q_{\theta_{ij}}(y_{ij}|\mathbf{x}_{ij};\psi)=y_{ij}$. If a hierarchical structure does explain part of the variability in the population data, units within clusters (areas) defined by this hierarchy are expected to have similar M-quantile coefficients. When the conditional M-quantiles are assumed to follow a linear model, with $\beta_\psi(q)$ a sufficiently smooth function of q, this suggests a predictor of $\bar Y_i$ of the form

$$\hat{\bar Y}_i^{MQ}=N_i^{-1}\left[\sum_{j\in s_i}y_{ij}+\sum_{j\in r_i}\mathbf{x}_{ij}^T\hat\beta_\psi(\hat\theta_i)+\frac{N_i-n_i}{n_i}\sum_{j\in s_i}\Big\{y_{ij}-\boldsymbol{x}_{ij}^T\hat\beta_\psi(\hat\theta_i)\Big\}\right] \tag{1}$$

(Tzavidis et al. 2010), where $\hat\theta_i=n_i^{-1}\sum\hat\theta_{ij}$ is an estimate of the average value of the M-quantile coefficients θ_{ij} for units in area i. These $\hat\theta_{ij}$ are obtained by solving $\hat Q_{\theta_{ij}}(y_{ij}|\mathbf{x}_{ij};\psi)=y_{ij}$ for θ_{ij} with $\hat Q_q$ denoting the estimated value of $Q_q(y_{ij}|\mathbf{x}_{ij};\psi)$ at q. For possible alternative choices of $\hat\theta_i$ see Chambers and Tzavidis (2006).

Tzavidis et al. (2010) refer to Expression (1) as the bias adjusted M-quantile predictor of $\bar Y_i$, derived as the mean functional of the Chambers and Dunstan (1986) estimator of the distribution function.

2.1. Failure of Benchmarking and Neutral Shrinkage Properties

The MQ predictors do not satisfy the benchmarking property. To see this, note first that $\hat{\bar Y}_i^{MQ}$ has an interesting GREG-like representation:

$$\begin{aligned}\hat{\bar Y}_i^{MQ}&=N_i^{-1}\sum_{j\in U_i}\mathbf{x}_{ij}^T\hat\beta_\psi(\hat\theta_i)+n_i^{-1}\left\{\sum_{j\in s_i}y_{ij}-\sum_{j\in s_i}\mathbf{x}_{ij}^T\hat\beta_\psi(\hat\theta_i)\right\}\\&=n_i^{-1}\sum_{j\in s_i}y_{ij}+\left\{N_i^{-1}\sum_{j\in U_i}\mathbf{x}_{ij}^T-n_i^{-1}\sum_{j\in s_i}\mathbf{x}_{ij}^T\right\}\hat\beta_\psi(\hat\theta_i)\end{aligned} \tag{2}$$

Using a more compact notation (2) may be rewritten as $\hat{\bar{Y}}_i^{MQ} = \hat{\bar{y}}_i + \left(\bar{\mathbf{X}}_i^T - \hat{\bar{\mathbf{x}}}_i^T\right)\hat{\beta}_\psi(\hat{\theta}_i)$ with $\bar{\mathbf{X}}_i^T = N_i^{-1}\sum_{j\in U_i}\mathbf{x}_{ij}^T$, $\hat{\bar{\mathbf{x}}}_i^T = n_i^{-1}\sum_{j\in s_i}\mathbf{x}_{ij}^T$ and $\hat{\bar{y}}_i = n_i^{-1}\sum_{j\in s_i} y_{ij}$

Let us assume, for simplicity, that we select a simple random sample from each area i and let $w_i = n_i/n$ be the sampling fraction in area i with $n = \sum_i n_i$. The direct estimator of the overall population mean $\bar{Y}$, $\hat{\bar{y}} = n^{-1}\sum_{i=1}^m\sum_{j=1}^{n_i} y_{ij}$ may be written as $\hat{\bar{y}} = n^{-1}\sum_{i=1}^m n_i\hat{\bar{y}}_i$, which is the weighted average of small area mean estimators. This property is desirable for all small area estimators, and it is, in this case, an alternative statement of the benchmarking property. When the small area mean is estimated by $\hat{\bar{Y}}_i^{MQ}$ we have that

$$n^{-1}\sum_{i=1}^m n_i\hat{\bar{Y}}_i^{MQ} = \hat{\bar{y}} + \sum_{i=1}^m w_i\left(\bar{\mathbf{X}}_i^T - \hat{\bar{\mathbf{x}}}_i^T\right)\hat{\beta}_\psi(\hat{\theta}_i) \tag{3}$$

from which it is clear that the estimator is not benchmarked to the overall direct estimator of the mean because, in general, $\sum_{i=1}^m d_i = \sum_{i=1}^m w_i\left(\bar{\mathbf{X}}_i^T - \hat{\bar{\mathbf{x}}}_i^T\right)\hat{\beta}_\psi(\hat{\theta}_i) \neq 0$. Futhermore, this property will not be satisfied for general sampling designs and weighted estimators of the overall mean $\bar{Y}$. About (3) we note that d_i will be small for large n_i; more precisely they will be $O_p(n_i^{-1/2})$ whenever $\left(\bar{\mathbf{X}}_i^T - \hat{\bar{\mathbf{x}}}_i^T\right) = O_p(n_i^{-1/2})$, $w_i = O_p(1)$ and $\hat{\beta}_\psi = O_p(1)$, where $O_p(\,)$ denotes the ordinary of convergence in probability. Of course, as we are focusing on small area estimation, large n_is are not of special interest here.

As far as neutral shrinkage is concerned, we note that, assuming a normal linear mixed model the set $\hat{\bar{Y}}_i^{MQ}$, $i = 1, \ldots, m$ is overdispersed with respect to the underlying population parameters:

$$\sum_{i=1}^m w_i\left(\hat{\bar{Y}}_i^{MQ} - \hat{\bar{Y}}_\cdot^{MQ}\right)^2 > \sum_{i=1}^m w_i\left(\bar{Y}_i - \bar{Y}\right)^2 \tag{4}$$

where $\hat{\bar{Y}}_\cdot^{MQ} = \sum_{i=1}^m w_i\hat{\bar{Y}}_i^{MQ}$, as will be confirmed by the simulation results of Section 4. The behaviour of M-quantile based predictors is then more similar to that of direct estimators and in contrast with that of the over-shrinking EB predictors (see EURAREA Consortium 2004 Section B.3). When outliers are present in the data and normality fails, the over-shrinkage of EB predictors becomes severe and also the ensemble of $\hat{\bar{Y}}_i^{MQ}$ exhibits a variance smaller than the actual set of population parameters. This effect, detected by the authors using simulations, will also be apparent when analysing the outlier affected data of Section 5.

3. Modified M-quantile Small Area Estimators

In this section we introduce adjusted MQ estimators. We consider two alternative approaches. The first is based on constraining M-quantile regression. It can be applied to obtain benchmarking MQ small area estimates, but cannot be easily extended to the correction of over/under-shrinkage as this would involve quadratic constraining that is very difficult to manage. The second approach is based on an ex-post "first two moments" matching procedure parallel to that commonly used to adjust the over-shrinkage of EB estimators (Rao 2003, Section 9.6). The two approaches may be integrated as the output of the benchmarking procedure may be used for the over/under shrinkage correction.

In this case the output of the second procedure would be an MQ estimator benchmarked and satisfying neutral shrinkage. In Section 3.1 we illustrate the method to obtain benchmarking MQ estimators (denoted BMQ); in Section 3.2 we describe the procedure to achieve neutral shrinkage.

3.1. Constrained M-estimation of Regression Parameters in Quantile Regression

The constrained robust regression model (Eddy and Kadane 1982) can be generalized to a model for the M-quantile of order q of the conditional distribution of y given x. Let $\mathbf{H}$ be a $h \times p$ matrix and suppose we want the vector $\beta_\psi(q)$ to match

$$\mathbf{H}\beta_\psi(q) = \mathbf{d} \tag{5}$$

where $\mathbf{d}$ is an $h \times 1$ vector of values ($\mathbf{d}$ may be a vector of zeroes). For specified q, a constrained estimate of the vector of the regression parameters $\beta_\psi(q)$ may then be obtained by minimizing

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} \rho_q\Big(y_{ij} - \mathbf{x}_{ij}^T\beta_\psi(q)\Big) + \Lambda^T(\mathbf{d} - \mathbf{H}\beta_\psi(q)) \tag{6}$$

where $\Lambda^T = (\lambda_1, \lambda_2, \ldots, \lambda_h)$ is a $1 \times h$ vector of Lagrange multipliers and $\rho_q(\cdot)$ is a loss function associated with the influence function $\psi_q(\cdot)$ introduced in Section 2. The constrained estimate of $\beta_\psi(q)$ may then be obtained by differentiating (6) with respect to $\beta_\psi(q)$ and Λ, setting the derivatives equal to zero and solving the normal equations

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} \psi_q\Big(y_{ij} - \mathbf{x}_{ij}^T\beta_\psi(q)\Big)\mathbf{x}_{ij} - \mathbf{H}^T\Lambda = 0 \tag{7}$$

Because (7) is a system of nonlinear equations, an iterative method is used for its solution. Let $w_\psi(r) = \psi_q(r)/r$ and $w_{\psi ij} = w_\psi(r_{ij})$, with $r_{ij} = y_{ij} - \mathbf{x}_{ij}^T\beta_\psi(q)$. Then (7) can be written as

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} w_{\psi ij}\Big(y_{ij} - \mathbf{x}_{ij}^T\beta_\psi(q)\Big)\mathbf{x}_{ij} - \mathbf{H}^T\Lambda = 0 \tag{8}$$

The steps of the iteratively reweighted least squares algorithm are as follows:

1. For specified q define an initial estimates $\beta_\psi^{(0)}(q)$.
2. At each iteration t, calculate the residuals $r_{ij}^{(t-1)} = y_{ij} - \mathbf{x}_{ij}^T\beta_\psi^{(t-1)}(q)$ and associated weights $w_{\psi ij}^{(t-1)}$ from the previous iteration.
3. Compute the new weighted least squares estimates subject to the constrained $\mathbf{H}\beta_\psi(q) = \mathbf{d}$:

$$\begin{aligned}\hat{\beta}_\psi^{(t)}(q) = {} & [\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}^T(\mathbf{H}\mathbf{A}^{-1}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{A}^{-1}]\mathbf{X}^T\mathbf{W}^{(t-1)}\mathbf{y} \\ & + \mathbf{A}^{-1}\mathbf{H}^T(\mathbf{H}\mathbf{A}^{-1}\mathbf{H}^T)^{-1}\mathbf{d}\end{aligned} \tag{9}$$

Here $\mathbf{X}$ is the matrix of order $n \times p$ of sample x values and $\mathbf{A} = \mathbf{X}^T\mathbf{W}^{(t-1)}\mathbf{X}$, $\mathbf{y}$ is the vector of n sample values for y. The matrix $\mathbf{W}^{(t-1)} = diag(w_{\psi ij})$ is a diagonal matrix of order n with entry corresponding to a particular sample observation set equal to the weight $w_{\psi ij}$.

4. Repeat Steps 1–3 until convergence. Convergence is achieved when the difference between the estimated model parameters obtained from two successive iterations is negligible. For details about the convergence of iterative reweighted least squares with constraints see Dempster et al. (1980).

The algorithm has been used in the simulation experiments (Section 4) and in the application (Section 5) and a satisfactory convergence has been obtained in 10 to 20 iterations. The R code that implements this algorithm is available from the authors.

The algorithm is applied to achieve benchmarking as follows. Assume simple random sampling within the areas and also that we are interested in the consistency of the estimated small area means with the estimate of the population mean using the overall sample. We have the following benchmarking equation:

$$\sum_{i=1}^{m} w_i \hat{\bar{Y}}_i^{MQ} = \sum_{i=1}^{m} w_i \left\{ \hat{\bar{y}} + \left(\bar{\mathbf{X}}_i^T - \hat{\bar{\mathbf{x}}}_i^T \right) \hat{\beta}_\psi(\hat{\theta}_i) \right\} = \hat{\bar{y}}$$

so, in view of (3) Equation (5) may be rewritten as:

$$\sum_{i=1}^{m} w_i \left(\bar{\mathbf{X}}_i^T - \hat{\bar{\mathbf{x}}}_i^T \right) \hat{\beta}_\psi(\hat{\theta}_i) = 0 \tag{10}$$

The benchmarking equation is expressed in terms of means, but it can be equivalently written in terms of totals. To derive the benchmarked MQ predictors, for short $\left\{ \hat{\bar{Y}}_i^{BMQ} \right\}$, we have to consider that the constraint (10) acts simultaneously on all the area-specific regression coefficients, so simultaneous constrained estimation of $\{\hat{\beta}_\psi(\hat{\theta}_i), \ldots, \hat{\beta}_\psi(\hat{\theta}_m)\}$ for the m small areas is required. The size of the vector of the M-quantile regression parameters becomes $(m \times p) \times 1$ and consequently the solution of the system of normal equations requires the inversion of a matrix of size $(n \times m) \times (m \times p)$. As a consequence, obtaining these constrained estimates may be computationally demanding in applications with a large number of areas.

Relaxing the assumption of simple random sampling and assuming more general sampling designs we will have that $\hat{\bar{Y}} = N^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} g_{ij} y_{ij}$ with $g_{ij} = \pi_{ij}^{-1}$ or defined in some more complex way but such that $\sum_{i=1}^{m} \sum_{j=1}^{n_i} g_{ij} = N$. A popular direct estimator is in this case given by $\hat{\bar{y}}_{iw} = g_i^{-1} \sum_{j=1}^{n_i} g_{ij} y_{ij}$ with $g_i = \sum_{j=1}^{n_i} g_{ij}$. As a consequence $\hat{\bar{y}} = 1/N \sum_{i=1}^{m} g_i \hat{\bar{y}}_{iw} = \sum_{i=1}^{m} w_i^* \hat{\bar{y}}_{iw}$ with $\sum_{i=1}^{m} w_i^* = 1$ as $\sum_{i=1}^{m} g_i = N$. The benchmarking constraint is then expressed by this equation

$$\sum_{i=1}^{m} w_i^{**} \hat{\bar{Y}}_i^{MQ} = \sum_{i=1}^{m} w_i^* \hat{\bar{y}}_{iw}$$

We may define $w_i^{**} = w_i^*$ even if formally this is not necessary. If we want to avoid the use of weights, we may keep $w_i^{**} = n_i/n$ or, if this information is available, $w_i^{**} = N_i/N$. Anyway, we will have that $\sum_{i=1}^{m} w_i^{**} \hat{\bar{y}}_i \neq \sum_{i=1}^{m} w_i^* \hat{\bar{y}}_{iw}$ so we need to modify (10) as follows:

$$\sum_{i=1}^{m} w_i^{**} \left(\bar{\mathbf{X}}_i^T - \hat{\bar{\mathbf{x}}}_i^T \right) \hat{\beta}_\psi(\hat{\theta}_i) = d$$

with $d = \sum_{i=1}^{m} w_i^* \hat{\bar{y}}_i - \sum_{i=1}^{m} w_i^{**} \hat{\bar{y}}_i$.

For the set $E = \{i|n_i = 0\}$ of the out of sample areas, i.e., areas where $n_i = 0$, consistently with Chambers and Tzavidis (2006), we may define $\hat{\bar{Y}}_i^{MQ} = \bar{\mathbf{X}}_i^T \hat{\beta}_\psi(0.5)$. In the benckmarking equation w_i^* cannot be used since $g_i = 0, i \in E$. In this case $w_i^{**} = N_i/N$ is a more sensible option.

3.2. Adjusted M-quantile Small Area Estimators With Neutral Shrinkage

In order to obtain a set of modified MQ predictors that satisfies benchmarking and neutral shrinkage (at least approximately), we propose a strategy that is similar to that of Rao (2003, Section 9.6). More specifically, given a set of predictors $\left\{\hat{\bar{Y}}_i^*\right\}_{1\le i\le m}$ of the small area means, we look for a new set of estimators $\{t_i\}_{1\le i\le m}$ that minimizes

$$\sum_{i=1}^{m} w_i\left(\hat{\bar{Y}}_i^* - t_1\right)^2$$

and satisfies benchmarking and neutral shrinkage, i.e., is subject to the constraints:

1. $\sum_{i=1}^m w_i t_i = c_1$;
2. $\sum_{i=1}^m w_i (t_i - t_\cdot)^2 = c_2$,

where $w_i = n_i/n$ or any other "weight" such that $\sum_{i=1}^m w_i = 1$, $t_\cdot = \sum_{i=1}^m w_i t_i$ and c_1 and c_2 are known constants. The constant c_1 will be a reliable estimator of the overall population mean, typically $\hat{\bar{y}}$ or some other (possibly survey weighted) model-free estimator.

The constant c_2, that can be rewritten as $c_2 = \sum_{i=1}^m w_i t_i^2 + c_1^2$, should be a suitable measure of the variance "between" the areas. Note that constraint 1 is redundant when the neutral shrinkage correction is applied to benchmarking estimators such as the BMQ of Section 3.1. For nonbenchmarking estimators our procedure may be seen as a crude way to attain the benchmarking property, that may be applied when the method of section 3.1 is impractical because of its computational complexity.

It may be shown (see the Appendix) that:

$$t_i^{opt} = c_1 + \left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}_\cdot^*\right)\sqrt{\frac{c_2}{\sum_{i=1}^m w_i\left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}_\cdot^*\right)^2}} \tag{11}$$

If we set $c_1 = \hat{\bar{Y}}_\cdot^*$ with $\hat{\bar{Y}}_\cdot^* = \sum_i w_i \hat{\bar{Y}}_i^*$ we will also have that

$$t_i^{opt} = \sqrt{\frac{c_2}{\sum_{i=1}^m w_i\left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}_\cdot^*\right)^2}}\hat{\bar{Y}}_i^* + \left(1 - \sqrt{\frac{c_2}{\sum_{i=1}^m w_i\left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}_\cdot^*\right)^2}}\right)\hat{\bar{Y}}_\cdot^*$$

In practice, in the calculation of (11) the problem is to find a reasonable value for c_2. Under a linear mixed model for the data, c_2 is a measure of the variation between unobservable area-specific model parameters that may be calculated using the estimates of the variance components. Without recourse to an explicit model, we define the "ideal" c_2 as the unweighted variance between the area-specific population means, i.e.,

$$c_2^* = \frac{1}{m-1}\sum_{i=1}^{m}\left(\bar{Y}_i - \bar{Y}\right)^2$$

To estimate this quantity we cannot use ordinary design-based estimators of $\bar{Y}_i$ and $\bar{Y}$ since they are known to be over-dispersed when the sample sizes in the areas are small (Fabrizi 2009).

Let's write $y_{ij} = \mathbf{x}_{ij}^T \beta_\psi(0.5) + e_{ij}$. It follows that:

$$\bar{Y}_i = N_i^{-1} \sum_{i=1}^{N_i} y_{ij} = \bar{\mathbf{X}}_i^T \beta_\psi(0.5) + h_i$$

The term $h_i = N_i^{-1} \sum_{j=1}^{N_i} e_{ij} = \bar{Y}_i - \bar{\mathbf{X}}_i^T \beta_\psi(0.5)$ may be thought of as a "pseudo random effect", since it captures the average deviation of units in the same area from the median regression plane. We have that $\bar{Y} = N^{-1} \sum_{i=1}^{m} N_i \bar{Y}_i = \bar{\mathbf{X}}^T \beta_\psi(0.5) + \bar{h}$ with $\bar{h} = N^{-1} \sum_{i=1}^{m} N_i h_i$ and, as a consequence

$$c_2^* \cong \beta_\psi(0.5)^T \mathbf{S}_{XX} \beta_\psi(0.5) + \frac{1}{m-1} \sum_{i=1}^{m} (h_i - \bar{h})^2$$

where

$$\mathbf{S}_{XX} = \frac{1}{m-1} \sum\nolimits_{i=1}^{m} (\bar{\mathbf{X}}_\mathbf{i} - \bar{\mathbf{X}})(\bar{\mathbf{X}}_\mathbf{i} - \bar{\mathbf{X}})^T.$$

The equality is only approximate since, in developing the square, we omit the double product term that can be shown to be negligible with respect to the two main addends.

We may estimate h_i with $\tilde{h}_i = n_i^{-1} \sum_{j=1}^{n_i} \left(y_{ij} - \mathbf{x}_{ij}^T \hat{\beta}_\psi(0.5) \right)$; the associated estimator of the second term in c_2^*, i.e., $(m-1)^{-1} \sum_{i=1}^{m} (\tilde{h}_i - \bar{\tilde{h}})^2$, with $\bar{\tilde{h}} = N^{-1} \sum_{i=1}^{m} N_i \tilde{h}_i$ is likely to be unstable and liable to the influence of outlying residuals. In line with the robust estimation approach adopted in this article we then propose the following estimator of c_2^*:

$$\hat{c}_2^* \cong \hat{\beta}_\psi(0.5)^T \mathbf{S}_{XX} \hat{\beta}_\psi(0.5) + \frac{1}{m-1} \sum_{i=1}^{m} [\psi(\tilde{h}_i - \bar{\tilde{h}})]^2 \tag{12}$$

where $\psi(\cdot)$ is an influence function such as the Huber proposal 2 already mentioned in Section 2. This function depends on a tuning constant c that should be chosen in accordance with the data at hand and the sizes of the area-specific samples on which the calculation of the $\tilde{h}_i$ is based. The more the occurrence of the outliers is likely and the smaller the area-specific sample sizes, the more pronounced the smoothing operated through $\psi(\cdot)$ is expected to be (i.e.,the smaller c should be).

We denote the estimators obtained following this procedure $\hat{\bar{Y}}_i^{CBMQ}$. In Sections 4 and 5 they will be compared to the estimators obtained constraining the Empirical Best $\left(\left\{ \hat{\bar{Y}}_i^{EB} \right\} \right)$ according to the procedure illustrated in Rao (2003, Section 9.6) and that will be denoted as $\left\{ \hat{\bar{Y}}_i^{CEB} \right\}$.

3.3. MSE Estimation of MQ Estimators

Mean Squared Error (MSE) estimation of M-quantile based small area mean estimators relies on the approach described in Chambers et al. (2008). Since the estimates $\hat{\beta}_\psi(q)$ of the M-quantile regression coefficients can be expressed as linear combinations of the sample

y values, it follows that, for fixed $\hat{\theta}_i$, the estimator of the area i, $\hat{\bar{Y}}_i^{MQ}$ can be written as linear combinations of these sample values; a first order approximation to its MSE can be developed using the arguments in Royall and Cumberland (1978). Let $\{b_{ij}; j \in s\}$ denote the set of weights that define each of the M-quantile predictors. This approach then leads to a MSE estimator of the form

$$mse\left(\hat{\bar{Y}}_i^{MQ}\right) = \frac{1}{N_i^2}\left[\sum_{j\in s_i}\left\{f_{ij}^2 + \frac{N_i - n_i}{n_i - 1}\right\}\left(y_{ij} - \mathbf{x}_{ij}^T\hat{\beta}_\psi(\hat{\theta}_i)\right)^2 + \sum_{j\in s\setminus s_i} f_{ij}^2\left(y_{ij} - \mathbf{x}_{ij}^T\hat{\beta}_\psi(\hat{\theta}_i)\right)^2\right] \tag{13}$$

with $f_{ij} = b_{ij} - 1$ if $j \in s_i$ and $f_{ij} = b_{ij}$ otherwise. Since $\hat{\bar{Y}}_i^{MQ}$ is an approximately unbiased estimator of the small area mean, the squared bias will not significantly impact the MSE. The main limitation of the MSE estimator is that it does not account for the variability introduced in estimating the area specific θ_i's.

The MSE estimator (14) can be used to formulate an estimator of the MSE of the constrained estimators. Following (Rao 2003, p. 279) a measure of uncertainty associated with $\hat{\bar{Y}}_i^{BMQ}$ and $\hat{\bar{Y}}_i^{CBMQ}$ can be obtained by

$$mse\left(\hat{\bar{Y}}_i^{BMQ/CBMQ}\right) = mse\left(\hat{\bar{Y}}_i^{MQ}\right) + \left(\hat{\bar{Y}}_i^{MQ} - \hat{\bar{Y}}_i^{BMQ-CBMQ}\right)^2 \tag{14}$$

This is a somewhat crude method and an empirical alternative to the analytical estimator of MSE may be represented by a bootstrap procedure. The definition and evaluation of such procedure is an object of our current research.

4. A Simulation Study

In this section we present a Monte Carlo study for checking whether the adjustment procedure illustrated in Section 3.2, including the proposed estimator for c_2, effectively works; we also aim at assessing the impact of the adjustment on the MSE and the bias of the predictors. To do this we consider a population generated according to a normal linear mixed model, for which we know that Empirical Best predictors are very efficient and CEB predictors correct for over-shrinkage. We can then use them as sound terms of comparison. We also investigate how the proposed MSE estimator (15) tracks the true MSE of the CBMQ estimator. We consider a model-based simulation in which properties of the traditional and proposed estimators are evaluated with respect to the process that generates the finite population from which the samples are drawn.

Let $\{\bar{Y}_i\}_{1\le i\le m}$ be the set of area-specific population means and $\left\{\hat{\bar{Y}}_i^*\right\}_{1\le i\le m}$ the corresponding predictors, with $* = DIR, EB, CEB, MQ, CBMQ$; $\hat{c}_2^*$ defined in (13) is used as a guess for c_2.

At each Monte Carlo iteration, the values of the study variable y defined on a finite population of size $N = k4,200$, where k is a positive integer number, are generated according to the following Battese-Harter-Fuller model:

$$y_{ij} = \mathbf{x}_{ij}^T\beta + v_i + \epsilon_{ij}$$

with $i = 1, \ldots, m = 36$, $j = 1, \ldots, N_i$ and N_i ranging from $k50$ to $k200$. To study the behaviour of the estimators for growing sample sizes in a framework consistent with the

ordinary asymptotic of finite populations (see Isaki and Fuller 1982) we consider sequences of sample and population sizes setting $k = 2, 3, 5, 10, 20$. The random components are drawn from the following normal distributions: $v_i \overset{ind}{\sim} N(0, \sigma_v = 16)$ and $\epsilon_{ij} \overset{ind}{\sim} N\left(0, \sigma_e^2 = 100\right)$. The auxiliary variables $\mathbf{x}_{ij}^T = (1, x_{ij})^T$ are generated only once and held fixed for all Monte Carlo iterations. In particular $x_{ij} = x_i + u_{ij}$ with $x_i \overset{ind}{\sim} N\left(\mu_x = 194, \sigma_x^2 = 2\right)$ and $u_{ij} \overset{ind}{\sim} N\left(0, \sigma_u^2 = 25\right)$. This population structure reflects a situation in favour of the application of linear mixed models as only a small part of the difference among area means is explained by the auxiliary information, that attributable to the unobservable random effect. This population is essentially the same as the one considered in Torabi et al. (2009).

From the above population a stratified simple random sample is drawn, with strata given by the areas. The allocation of the sample is assumed to be proportional to the population size. Different total sample sizes with $n = k84$ are considered; as a result of proportional allocation area-specific sample sizes range from $k1$ to $k4$.

The predictors are compared in terms of:

1. their ability to estimate the actual descriptive variance of the "ensemble" of the area means, i.e., calculating the ratio between the variance of the "ensemble" of the estimates, and the same quantity defined on the set of the underlying population parameters:

$$AVR[\{\hat{\bar{Y}}^*\}] = R^{-1}\sum_{r=1}^{R} \frac{(m-1)^{-1}\sum_{i=1}^{m}\left(\hat{\bar{Y}}_{i,r}^* - \hat{\bar{Y}}_r\right)^2}{(m-1)^{-1}\sum_{i=1}^{m}\left(\bar{Y}_{i,r} - \bar{Y}_r\right)^2} \tag{15}$$

2. their average bias:

$$AB[\{\hat{\bar{Y}}^*\}] = R^{-1}\sum_{r=1}^{R} m^{-1}\sum_{i=1}^{m}\left(\hat{\bar{Y}}_{i,r}^* - \bar{Y}_{i,r}\right) \tag{16}$$

3. their average MSE:

$$AMSE[\{\hat{\bar{Y}}^*\}] = R^{-1}\sum_{r=1}^{R} m^{-1}\sum_{i=1}^{m}\left(\hat{\bar{Y}}_{i,r}^* - \bar{Y}_r\right)^2 \tag{17}$$

The index r is the counter of Monte Carlo replications, whose total number, R, is set equal to 5,000. Note that 2. and 3. are properties evaluated "on average" with respect to the set of the small area being studied. Average evaluation of small area estimators is in line with many simulation exercises in the current literature (see Rao 2003, Section 7.2.6).

The results of the simulation study are illustrated in Table 1 and Table 2. In Table 1 the quantities reported within parentheses for the CBMQ estimators are those calculated using the same guess for c_2 as in the case of CEB.

Focusing on the ability of the predictors to achieve neutral shrinkage, we have that both EB and MQ estimators converge to the value 1 of AVR as the average area-specific sample sizes grow large, but from different sides. As expected, EB predictors "over-shrink" the distribution; the estimates based on the quantile method are more dispersed than the actual parameters, although not as much as the area-specific sample means. This overdispersion may be attributed to the poor estimation of θ_i when area-specific samples are small; it decreases as k grows; this should be expected in view of (1).

Table 1. Model-based simulation results. Within parentheses the values of AVR, AB and AMSE computed using the same values for c_2 used for the CEB estimators

k	n	$\{\bar{Y}_i^*\}$	AVR	AB	AMSE
2	168	DIR	2.57	−0.08	57.22
2	168	EB	0.72	0.00	11.48
2	168	CEB	1.06	0.10	13.45
2	168	MQ	1.77	0.00	28.29
2	168	CBMQ	1.29 (1.06)	−0.09 (−0.21)	22.55 (19.81)
3	252	DIR	2.12	−0.01	38.38
3	252	EB	0.76	−0.01	9.07
3	252	CEB	1.06	0.08	10.44
3	252	MQ	1.55	0.00	19.00
3	252	CBMQ	1.14 (1.06)	−0.10 (−0.23)	15.28 (14.60)
5	420	DIR	1.70	0.00	23.11
5	420	EB	0.82	−0.01	6.64
5	420	CEB	1.05	0.05	7.49
5	420	MQ	1.36	−0.01	11.37
5	420	CBMQ	1.09 (1.04)	−0.06 (−0.06)	10.19 (9.56)
10	840	DIR	1.34	0.00	11.51
10	840	EB	0.88	0.00	4.14
10	840	CEB	1.04	0.06	4.73
10	840	MQ	1.17	0.00	5.70
10	840	CBMQ	1.04 (1.04)	−0.03 (−0.03)	5.61 (5.70)
20	1,680	DIR	1.17	0.00	5.69
20	1,680	EB	0.94	0.00	2.38
20	1,680	CEB	1.03	0.03	2.80
20	1,680	MQ	1.09	0.00	2.90
20	1,680	CBMQ	1.01 (1.03)	0.00 (−0.01)	3.05 (3.23)

Comparing the CEB and CBMQ estimates, the performances in terms of AVR are really close whenever the same value of c_2 is plugged into the constraining procedure. The use of $\hat{c}_2^*$ is not as effective when k is small even if the impact goes in the right direction in all the cases; anyway it should be noted that the simulation is conducted under the assumption of the normal linear mixed model which favours the correction incorporated into the CEB estimator.

The bias is moderate in all cases, with the one exception of the CBMQ predictor when $k = 2$ and $k = 3$. This implies that the distributions of the "ensembles" of the estimates based on both the EB and the MQ methods are centered about their true means.

As regards the efficiency measured by AMSE, EB predictors are far more efficient than MQ estimators when k is small. For larger sample sizes this difference dwindles. This confirms the expectation that when the assumptions of EB predictors hold, as in this simulation, they yield big gains in efficiency, especially when area-specific sample sizes are very small. Methods based on robust modeling and weaker assumptions, such as the M-quantile, become valid alternatives when more than a few units are observed in each area.

Table 2. Across areas distribution of true (i.e., Monte Carlo) root mean squared errors (True RMSE) and area averages of estimated root mean squared errors (Est. RMSE)

		Percentile of across areas distribution					
k	Indicator	10	25	Median	Mean	75	90
2	True RMSE	2.948	3.617	4.432	4.594	6.155	6.281
2	Est. RMSE	3.251	3.969	4.660	4.603	5.484	5.686
3	True RMSE	2.405	2.970	3.638	3.786	5.034	5.153
3	Est. RMSE	2.731	3.366	3.990	4.041	5.072	5.151
5	True RMSE	1.935	2.437	2.979	3.094	4.116	4.214
5	Est. RMSE	2.079	2.648	3.183	3.256	4.196	4.281
10	True RMSE	1.463	1.822	2.199	2.295	3.038	3.150
10	Est. RMSE	1.538	1.906	2.305	2.386	3.109	3.185
20	True RMSE	1.072	1.340	1.616	1.695	2.228	2.292
20	Est. RMSE	1.122	1.381	1.678	1.744	2.267	2.326

Moreover the CBMQ predictors, less dispersed than their unconstrained counterparts, show a lower AMSE. The gain in efficiency is not very big and depends on the amount of under-shrinkage effectively corrected; for this reason it reduces when k grows. In any case, this is in sharp contrast with the behaviour of the constrained EB predictors that are, by construction, suboptimal in terms of MSE with respect to unconstrained EB predictors. Intuitively, if we consider (11), we may note that, because of the underdispersion of the MQ predictors, the quantity under the square root is less than 1, thus leading to a variance reduction that over-compensates the increase in the bias. As a consequence of this fact the gap between constrained MQ and EB predictors is smaller than that between ordinary (unconstrained) MQ and EB predictors.

To sum up, we found that M-quantile estimators are comparable to EB predictors even when data are generated under a normal linear mixed model, provided that the area-specific sample sizes are not too small. The over-shrinkage correction illustrated in the previous section is effective under the same circumstances.

Table 2 shows key percentiles of the across area distributions of the area level true and estimated root mean squared errors (the latter based on (14) and averaged over the simulations) of the CBMQ predictor. In general the proposed MSE estimator (14) provides a good approximation to the true MSE.

We also run simulations in which $\epsilon_{ij} \overset{ind.}{\sim} 10/\sqrt{3}t_3$ (a re-scaled Student's t with 3 degrees of freedom) instead of $\epsilon_{ij} \overset{ind.}{\sim} N(0, 100)$ to check the robustness of the introduced constraining methods to the presence of outliers. The performances of CBMQ and its MSE estimator are still good. Anyway, about the latter there emerges a tendency to overestimate when the average area-specific samples are very small (i.e., very small k). Detailed results are available from the authors upon request.

We note that the aim of the simulation exercise of this section was not to compare the efficiency of the various methods when outliers are present (concerning this, see Salvati et al. 2011) as we may expect that M-quantile regression based methods do better than Empirical Best predictors in this case, but to compare the effectiveness of the adjustments needed to achieve neutral shrinkage and their impact in terms of efficiency.

5. An Application to Survey and Satellite Data

Battese et al. (1988) analyse survey and satellite data for corn and soybean in a part of Iowa. Their objective is to predict the means of the areas under corn and soybean for twelve counties (small areas) in North Central Iowa.

Data are from the 1978 June Enumerative Survey and include information on corn and soybean areas at individual pixel and segment level. The data set contains the number of segments in each county, the number of hectares of corn and soybean for each sample segment, and the number of pixels per segment in each county classified as corn and soybean. The linear mixed model used for the small area mean hectares of corn and soybean per segment is described in the paper by Battese et al. (1988). We use this well-known data set to compare the performances of EB, MQ predictors and their adjusted versions aimed at achieving neutral shrinkage (CEB, CBMQ). The analysis of the soybean variable represents a situation in which normality approximately holds, while for corn there is an influential outlier (in Hardin county). The presence of an outlier in the data is a typical departure from normality. We analyse the corn variable with and without the Hardin county outlier, to evaluate its impact on the behaviour of the "ensemble" predictors.

Table 3 presents the EB, CEB, MQ, CBMQ estimates for the mean hectares of corn and soybean per segment for each county. The procedure described in Section 3.2 involves the unknown quantities c_1 and c_2. The first constant c_1 is equal to the direct estimate of the mean hectares of crop (soybean) obtained for the union of the twelve counties, the second constant c_2 is estimated by (13) for the adjusting of MQ estimators and according to the method described in Rao (2003, Section 9.6) for the EB predictors. The averages and the between-areas variances of the small area estimates are reported in the last two rows of Table 3. For CEB and CBMQ the results coincide with the c_1 and c_2 constraints.

The small area predictions from all these methods are somewhat similar for most counties with a few exceptions. As may be expected, differences are larger for areas with one or very few observations, smaller for the counties with somewhat larger sample sizes. Moreover, we note that constraining the EB and MQ area predictors to known values for planned domains reduces the influence of the outlying observation in both cases. In fact the CEB and CBMQ predictors are similar for all the counties, Hardin included.

Under normality – soybean data – our results are in line with the results of the simulation experiment in Section 4. The MQ estimator has a between-areas variance equal to 380.32 against a $\hat{c}_2^*$ of 335.36. The EB predictors tend to over-shrink the distribution (307.81) and the CEB corrects this behaviour ($\hat{c}_2 = 339.22$).

As regards corn data, when the outlier is included EB predictors over-shrink the distribution dramatically, but they are far less shrunken when the outlier is removed. The estimates of c_2 with and without the outlier are widely different. The reason for this lies in the big impact that an outlier has on the estimation of variance components. MQ predictors, although influenced to some extent by the presence of the outlier, are more robust in this respect; in particular the estimates of c_2 based on (12) with and without the outlier are reasonably close. In a situation where normality does not hold, MQ estimators, differently from the normal linear mixed model setting of previous sections, tend to over-shrink the actual variance of area specific population parameters. So, differently from the EB predictors that always over-shrink, the MQ predictors may over- or under-shrink depending

Table 3. Predicted mean hectares of soybean and corn per segment

		Soybean				Corn				Corn with outliers			
County	n_i	EB	CEB	MQ	CBMQ	EB	CEB	MQ	CBMQ	EB	CEB	MQ	CBMQ
Cerro Gordo	1	78.6	76.3	74.0	73.8	122.3	122.0	127.8	126.9	122.2	122.2	129.7	130.2
Hamilton	1	94.4	92.9	100.8	99.0	126.1	126.1	133.2	132.1	123.2	123.5	133.8	134.7
Worth	1	87.3	85.5	80.7	80.2	107.4	105.4	93.0	92.8	114.0	112.2	84.1	79.9
Humboldt	2	81.2	79.0	82.1	81.4	108.9	107.0	109.0	108.5	115.5	114.0	110.5	109.0
Franklin	3	66.2	63.3	62.8	63.4	143.9	146.0	149.5	148.1	135.9	139.2	149.5	152.1
Pocahontas	3	113.8	113.2	113.4	110.9	111.9	110.3	116.7	116.0	108.3	105.2	116.9	116.1
Winnebago	3	97.8	96.4	101.5	99.7	113.6	112.2	110.9	110.3	117.1	116.1	112.4	111.1
Wright	3	112.3	111.7	113.6	111.1	122.0	121.6	123.6	122.8	122.6	122.8	124.0	123.9
Webster	4	109.8	109.0	109.3	107.0	115.1	113.9	117.6	116.9	110.9	108.4	117.4	116.7
Hancock	5	100.6	99.4	102.5	100.6	124.5	124.4	122.1	121.3	124.5	125.1	120.8	120.4
Kossuth	5	119.0	118.7	121.8	118.7	107.2	105.2	104.8	104.4	113.6	111.7	105.9	103.9
Hardin	5 (6)	74.8	72.3	71.8	71.8	142.8	144.7	143.0	141.8	131.3	133.5	131.5	132.1
$m^{-1}\sum \hat{\bar{Y}}_i^*$		96.74	95.34	97.52	95.34	121.68	121.26	120.93	121.26	120.65	120.38	120.81	120.38
$(m-1)^{-1}\sum\left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}^*\right)^2$		307.81	339.22	380.32	335.36	166.86	206.74	227.50	208.34	73.07	110.35	172.84	210.99

on the actual distribution of the data. In other cases we have found that they tend to over-shrink whenever the study variable has a skewed distribution.

6. Conclusions

The MQ based estimators are a recent and promising proposal in the small area literature and the analysis of their properties is an area of active research. In this article we explored the behaviour of MQ predictors with respect to two coherence properties, benchmarking and neutral shrinkage that are of interest to final users of small area estimates. Since the estimators introduced in Chambers and Tzavidis (2006) do not satisfy these properties, we proposed modified estimators.

As regards benchmarking, our solution is consistent with the M-quantile regression framework, thus it is theoretically more interesting than a simple ratio adjustment. It should be noted that obtaining these benchmarked MQ estimators may be computationally demanding when the (overall) sample size and the number of areas is large.

With respect to neutral shrinkage we found that the MQ estimators may under-shrink (under normality) or over-shrink (when the distribution of actual small area parameters is skewed); this behaviour is different from that of EB predictors, which always over-shrink. The solution proposed to obtain MQ predictors adjusted in this sense suffers from some limitations. Similarly to what is usually done for EB predictors, the correction is based on the first two moments; so the focus is mainly on normal or close to normal situations, and may not be sensible when the distribution of actual small area parameters is very skewed. Anyway, we keep the correction based on the first two moments as it represents the standard in the literature and in practical applications.

A possible area of future research is represented by the consideration of other coherence properties that are desirable to users, especially in official statistical agencies, and design consistency of the M-quantile predictors in particular the specifying also their asymptotic behaviour and the role of sampling weights.

Appendix

Proof of (11)

First note that

$$\sum_{i=1}^{m} w_i(t_i - t_.)^2 = \sum_{i=1}^{m} w_i t_i^2 + t_.^2 \sum_{i=1}^{m} w_i - 2t_. \sum_{i=1}^{m} w_i t_i = \sum_{i=1}^{m} w_i t_i^2 - t^2$$

We want to minimize

$$\phi = \sum_{i=1}^{m} w_i \left(\hat{\bar{Y}}_i^* - t_i \right)^2 - a_1 \left(\sum_{i=1}^{m} w_i t_i - c_1 \right) - a_2 \left(\sum_{i=1}^{m} w_i t_i^2 - c_2 - c_1^2 \right)$$

The first partial derivative in t_i is:

$$\frac{\partial \phi}{\partial t_i} = -2w_i \left(\hat{\bar{Y}}_i^* - t_i \right) - a_1 w_i - 2a_2 w_i t_i$$

Equating this derivative to 0 and solving for t_i we obtain

$$t_i^{opt} = \frac{1}{1-a_2}\left(\hat{\bar{Y}}_i^* + \frac{a_1}{2}\right)$$

Imposing the first constraint $\sum_{i=1}^m w_i t_i = c_1$ and solving in $a_1/2$ we obtain

$$\frac{a_1}{2} = c_1(1-a_2) - \hat{\bar{Y}}_.^*$$

leading to

$$t_i^{opt} = \frac{1}{1-a_2}\left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}_.^*\right) + c_1$$

Imposing the second constraint $\sum_{i=1}^m w_i(t_i - t_.)^2 = c_2$ and solving in $1/(1-a_2)$ we obtain

$$\frac{1}{1-a_2} = \sqrt{\frac{c_2}{\sum_{i=1}^m w_i\left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}_.^*\right)^2}}$$

$$t_i^{opt} = \sqrt{\frac{c_2}{\sum_{i=1}^m w_i\left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}_.^*\right)^2}}\left(\hat{\bar{Y}}_i^* - \hat{\bar{Y}}_.^*\right) + c_1$$

7. References

Battese, G., Harter, R., and Fuller, W. (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. Journal of the American Statistical Association, 83, 28–36.

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

Chambers, R. and Dunstan, R. (1986). Estimating Distribution Function from Survey Data. Biometrika, 73, 597–604.

Chambers, R. and Tzavidis, N. (2006). M-quantile Models for Small Area Estimation. Biometrika, 93, 255–268.

Chambers, R., Chandra, H., and Tzavidis, N. (2008). On Bias-Robust Mean Squared Error Estimation for Linear Predictors for Domains. Working Papers, 09-08 Centre for Statistical and Survey Methodology, The University of Wollongong, Australia, (Available from: http://cssm.uow.edu.au/publications).

Dempster, A., Laird, N., and Rubin, D. (1980). Iteratively Reweighted Least Squares for Linear Regression When Errors are Normal/Independent Distributed. Multivariate Analysis V, (ed.) P.R. Krishnaiah. Amsterdam: North–Holland, 35–57.

Eddy, W. and Kadane, J. (1982). The Cost of Drilling for Oil and Gas: An Application of Constrained Robust Regression. Journal of the American Statistical Association, 77, 262–269.

EURAREA Consortium. (2004). Enhancing Small Area Estimation Techniques to meet European Needs, Project Reference Volume. Available at http://www.statistics.gov.uk/eurarea/download.asp.

Fabrizi, E. (2009). A Comparison of Adjusted Bayes Estimators of an Ensemble of Small Area Parameters. Statistica, LXIX, 269–284.

Frey, J. and Cressie, N. (2003). Some Results on Constrained Bayes Estimators. Statistics and Probability Letters, 65, 389–399.

Ghosh, M. (1992). Constrained Bayes Estimation with Applications. Journal of the American Statistical Association, 87, 533–540.

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

Isaki, C. and Fuller, W.A. (1982). Survey Design under the Regression Superpopulation Model. Journal of the American Statistical Association, 77, 89–96.

Jiang, J. and Lahiri, P. (2006). Mixed Model Prediction and Small Area Estimation. Test, 15, 1–96.

Judkins, D. and Liu, J. (2000). Correcting the Bias in the Range of a Statistic Across Small Areas. Journal of Official Statistics, 16, 1–13.

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

Koenker, R. and D'Orey, V. (1987). Computing Regression Quantiles. Biometrika, 93, 255–268.

Rao, J.N.K. (2003). Small Area Estimation: John Wiley & Sons.

Royall, R. and Cumberland, W. (1978). Variance Estimation in Finite Population Sampling. Journal of the American Statistical Association, 73, 351–358.

Salvati, N., Ranalli, M., and Pratesi, M. (2011). Small Area Estimation of the Mean Using Non-parametric M-quantile Regression: A Comparison when a Linear Mixed Model Does Not Hold. Journal of Statistical Computation and Simulation, 81, 945–964.

Sinha, S. and Rao, J. (2009). Robust Small Area Estimation. Canadian Journal of Statistics, 37, 381–399.

Torabi, M., Datta, G., and Rao, J. (2009). Empirical Bayes Estimation under a Nested Error Linear Regression Model with Measurement Errors in the Covariates. Scandinavian Journal of Statistics, 36, 355–369.

Tzavidis, N., Marchetti, S., and Chambers, R. (2010). Robust Prediction of Small Area Means and Distributions. Australian and New Zealand Journal of Statistics, 52, 167–186.

Ugarte, M.D., Militino, A., and Goicoa, T. (2009). Benchmarked Estimates in Small Areas Using Linear Mixed Models with Restrictions. Test, 18, 342–364.

Received November 2009
Revised September 2011