Small Area Quantile Estimation

Jiahua Chen and Yukun Liu

University of British Columbia and East China Normal University

Abstract

Sample surveys are widely used to obtain information about totals, means, medians, and other parameters of finite populations. In many applications, similar information is desired for subpopulations such as individuals in specific geographic areas and socio-demographic groups. Often, the surveys are conducted at national or similarly high levels. The random nature of the probability sampling can result in few sampling units from many subpopulations that are not considered at the design stage. It is difficult to estimate the parameters of these subpopulations (small areas) with satisfactory precision and to evaluate the accuracy of the estimates. In the absence of direct information, statisticians resort to pooling information across small areas via suitable model assumptions and administrative archives and census data. In this paper, we propose three estimators of small area quantiles for populations admitting a linear structure with normal error distributions or error distributions satisfying a semiparametric density ratio model (DRM). We study the asymptotic properties of the DRM-based method and find it to be root-n consistent. Extensive simulation studies reveal the properties of the three methods under various possible populations. The DRM-based method is found to be significantly more efficient when the error distribution is skewed; otherwise, its efficiency is comparable to that of the other methods.

1 Introduction

Sample surveys are widely used to obtain information about totals, means, medians, and other parameters of finite populations. In many applications, similar information is desired for subpopulations such as individuals in specific geographic areas and socio-demographic groups. The estimation of finite subpopulation parameters is referred to as the small area estimation problem (Rao 2003). While the geographic areas may not be small, there is often a shortage of direct information for individual areas. Often, the surveys are conducted at national or similarly high levels. The random nature of the probability sampling can result in few sampling units from many subpopulations that are not considered at the design stage. It is difficult to estimate the parameters of these subpopulations with satisfactory precision and to evaluate the accuracy of the estimates.

Because of the scarcity of direct information from small areas, reliable estimates are possible only if indirect information from other areas is available and effectively utilized. This leads to a common thread of "borrowing strength." Statisticians also seek auxiliary information from sources such as administrative archives and census data to obtain an indirect estimate for the subpopulation parameter. This estimate may then be combined "optimally" with the direct estimate if available.

Pioneering work on small area estimation includes Fay and Herriot (1979), Prasad and Rao (1990), and Lahiri and Rao (1995). Research in this area has received increasing attention from both the public and private sectors (Fay and Herriot 1979; Schaible 1993; Kriegler and Berk 2010). The number of publications on this topic is increasing (Pfeffermann 2002, 2013; Jiang and Lahiri 2006; Ghosh et al. 2008; Jiang et al. 2010; Jiongo et al. 2013). Most studies focus on estimating small area means.

In sample surveys, the population distribution and quantiles are also important parameters of interest. There are many papers devoted to their efficient estimation in various situations, such as Chambers and Dunstan (1986), Francisco and Fuller (1991), Wang and Dorfman (1996), and Chen and Wu (2002). Recently, small area quantile estimations have also drawn substantial attention; see Tzavidis and Chambers (2005), Chambers and Tzavidis (2006), Molina and Rao (2010), and Chaudhuri and Ghosh (2011). A more detailed review will be given in Section 2.

In this paper, we propose three estimators of small area quantiles for populations admitting a linear structure with normal error distributions or error distributions satisfying a semiparametric density ratio model (DRM). In Section 3, we motivate and develop the new methods. In Section 4, we present theoretical properties of the DRM-based quantile estimators. They are found to be root-n consistent under some generic conditions; the technical proofs are given in the supplementary material. In Section 5, extensive simulation studies reveal properties of the three methods for various possible populations. The DRM-based method is found to be significantly more efficient when the error distribution is skewed; otherwise, its efficiency is comparable to that of the other methods. We end the paper with a summary and discussion.

2 Literature review

The nested-error (unit level) regression model (NER) of Battese, Harter, and Fuller (1988) has been widely adopted in the literature for small area estimation. Consider the situation where the population is composed of $m+1$ small areas, and n_k sampling units are obtained from the kth area ($k = 0, 1, 2, \dots, m$). Under this model, the univariate response value and its vector covariates on these sampling units satisfy

$$y_{kj} = \mathbf{x}_{kj}^{\tau}\boldsymbol{\beta} + v_k + \varepsilon_{kj}, \tag{1}$$

where v_k denotes an area-specific random effect and ε_{kj} is a random error. The homogeneous NER model assumption includes $v_k \sim N(0, \sigma_b^2)$, $\varepsilon_{kj} \sim N(0, \sigma^2)$, and that they are independent of each other and the covariates $\mathbf{x}_{kj}$. The assumption $\varepsilon_{kj} \sim N(0, \sigma^2)$ can be relaxed by allowing an area-specific σ^2 (heterogeneous NER or HNER) or replacing the normality by a semiparametric setting.

Under this model, the d_1-variate regression coefficient $\boldsymbol{\beta}$ is common across the small areas. Hence, samples from all the areas contain its information, and they are pooled to estimate $\boldsymbol{\beta}$. When the overall sample size $n = \sum_{k=0}^{m} n_k$ is large, its estimator $\hat{\boldsymbol{\beta}}$ has high precision. Suppose the population covariate means $\bar{\mathbf{X}}_k$ are known from, say, administrative records. Sensible indirect estimates of the population means $\bar{\mathbf{Y}}_k$ (of Y) would be $\hat{\bar{\mathbf{Y}}}_k = \bar{\mathbf{X}}_k^{\tau}\hat{\boldsymbol{\beta}}$. Direct estimates of $\bar{\mathbf{Y}}_k$, such as the regression estimator $\bar{y}_k + (\bar{\mathbf{X}}_k - \bar{\mathbf{x}}_k)\hat{\boldsymbol{\beta}}$ in obvious notation, can be combined to improve the efficiency. The specifics will be given later.

The above model places assumptions on y_{kj}. Another commonly used model (Fay and Herriot 1979) places assumptions on the area-level estimators in the form of $\hat{\bar{Y}}_k = \sum_j w_{kj} y_{kj} / \sum_j w_{kj}$ incorporating the survey design. At this stage, we downplay the importance of the design and do not explain the design weights w_{kj} and other issues. They will be part of our future development.

In some applications, conditional quantiles rather than expectations of Y given $\mathbf{x}$ are of interest. The following quantile regression model is a useful platform:

$$P(Y \leqslant \mathbf{x}^{\tau}\boldsymbol{\beta}_q | \mathbf{X} = \mathbf{x}) = q$$

for each $q \in (0, 1)$. Clearly, $\mathbf{x}^{\tau}\boldsymbol{\beta}_q$ provides another useful way to characterize the relationship. To save space, we cite only the ground-breaking paper Koenker and Bassett (1978) from an extensive literature on regression quantiles.

The regression quantile function $\mathbf{x}^{\tau}\beta$ may be regarded as a solution to $\min \mathbb{E}\{\rho_q(Y - \mathbf{X}^{\tau}\boldsymbol{\beta})|\mathbf{X}\}$ with a specific M-function $\rho_q(\cdot)$. Additional considerations lead to the use of a generic $\rho_q(\cdot)$ and hence the M-quantiles proposed by Breckling and Chambers (1988). Chambers and Tzavidis (2006) further extended the use of M-quantile models to small area estimation. In general, the qth M-quantile of the conditional distribution of Y is denoted $\mathbf{x}^{\tau}\boldsymbol{\beta}_{\psi}(q)$, where the subscript ψ denotes the specific M-function in the definition.

In the context of small area estimation, each unit with value $\mathbf{x}_j, y_j$ has a q_j value such that $y_j =$

$\mathbf{x}_j^\tau \boldsymbol{\beta}_\psi(q_j)$. Let the average q_j value over small area k be θ_k. Chambers and Tzavidis (2006) suggested that θ_k reflects the random fluctuation of small area k. Hence, the cumulative distribution function (cdf) of y of small area k may be estimated by

$$\hat{F}_k(t) = N_j^{-1}\big[\sum_{j\in s_k} I(y_{kj} \leqslant t) + \sum_{j\in r_k} I(\mathbf{x}_{kj}\boldsymbol{\beta}_\psi(\theta_j) \leqslant t)\big]$$

where s_k and r_k are sets of observed and unobserved units in small area k. The unknown $\boldsymbol{\beta}_\psi(\cdot)$ is fitted over the whole data set. The small area quantiles are estimated accordingly.

The M-quantile-based approaches have been successfully employed in applications; see Tzavidis and Chambers (2005) and Tzavidis et al. (2007). At the same time, they have some obvious limitations. First, one must have all the values of $\mathbf{x}$ in order to compute $\hat{F}_k$. Second, the estimation is done by predicting all the unobserved y values in the population. The empirical cdf based on the predicted values can be inconsistent if the lost randomness in the prediction is not restored (Chen, Rao, and Sitter, 2000). We are curious about the conditions under which the M-quantile-based quantile estimators are consistent.

Molina and Rao (2010) and Chaudhuri and Ghosh (2011) are two other important developments in quantile estimation. Molina and Rao (2010) postulated a parametric joint distribution of $\mathbf{y}_s$ and $\mathbf{y}_r$ (or the transformed response) where s and r stand for sets of observed and unobserved units in the population. Once the joint distribution is estimated optimally, the conditional distribution of $\mathbf{y}_r$ given $\mathbf{y}_s$ becomes available. The authors suggested sampling from this distribution to make up the unobserved $\mathbf{y}_r$. The approach works well for small sample means and the cumulative distribution function if we regard $I(y_{kj} \leqslant t)$ as a transformed response. The quantile estimation is a byproduct. Chaudhuri and Ghosh (2011) proposed a method that contains a substantial nonparametric component. However, this component is only for the posterior distribution of the parameters in a full parametric model on y. The posterior quantiles of y in small areas are used as estimates. Because of this, their method is in fact fully parametric.

3 Proposed methods

Model (1) has a built-in mechanism for small area quantile estimation. Let G_k be the cumulative distribution function of ε_{kj}. It can be seen that

$$\begin{aligned} P(y_{kj} \leqslant y) &= \mathbb{E}\{P(\varepsilon_{kj} \leqslant y - \nu_k - x_{kj}\beta\}|\nu_k, x_{kj}\} \\ &= \mathbb{E}\{G_k(y - \nu_k - x_{kj}\beta)\}. \end{aligned}$$

Hence, the population distribution of this small area is given by

$$F_k(y) = N_k^{-1} \sum_{j=1}^{N_k} G_k(y - \nu_k - x_{kj}\beta)$$

where N_k denotes the area population size.

For any $\alpha \in (0, 1)$, define the α-quantile of F_k as $\xi_k = \xi_{k,\alpha} = \inf\{y : F_k(y) \geqslant \alpha\}$. Let $\hat{F}_k(y)$ be an estimate of $F_k(y)$. The corresponding small area quantile is estimated as

$$\hat{\xi}_k = \hat{\xi}_{k,\alpha} = \inf\{y : \hat{F}_k(y) \geqslant \alpha\}. \tag{2}$$

Therefore, the small area quantile estimation problem becomes a cdf estimation problem.

3.1 Estimation under normality

Let $\tilde{\sigma}^2$, $\tilde{\sigma}_b^2$, and $\tilde{\boldsymbol{\beta}}$ be the MLEs of σ^2, σ_b^2, and $\boldsymbol{\beta}$ under the assumption that the error distributions are normal with equal variance across the small areas. Denote $\tilde{\gamma}_k = n_k\tilde{\sigma}_b^2/(\tilde{\sigma}^2 + n_k\tilde{\sigma}_b^2)$. An empirical best linear unbiased prediction (EBLUP) for the small area mean is given by

$$\tilde{\bar{Y}}_k = \bar{\mathbf{X}}_k^\tau\tilde{\boldsymbol{\beta}} + \tilde{\gamma}_k(\bar{y}_k - \bar{\mathbf{x}}_k^\tau\tilde{\boldsymbol{\beta}}) = \bar{\mathbf{X}}_k^\tau\tilde{\boldsymbol{\beta}} + \tilde{\gamma}_k\hat{\nu}_k. \tag{3}$$

Remark: because $\tilde{\boldsymbol{\beta}}$ is MLE, it is not strictly unbiased but the terminology (E)BLUP sticks. Note the shrinkage factor $\tilde{\gamma}_k$ for the random effect ν_k in the small area mean estimation. Let $\Phi(\cdot)$ be the cdf of the standard normal. Substituting $\Phi(\cdot/\sigma)$ for $G_k(\cdot)$ and so on in $F_k(y)$ leads to

$$F_k(y) = N_k^{-1} \sum_{j=1}^{N_k} \Phi\big(\{y - (\mathbf{x}_{kj} - \bar{\mathbf{X}}_k)^\tau \tilde{\boldsymbol{\beta}} - \bar{\mathbf{y}}_k\}/\tilde{\sigma}\big).$$

Its sample version, taking (3) into consideration, is our first cdf estimator:

$$\tilde{F}_k(y) = \frac{1}{n_k} \sum_{j=1}^{n_k} \Phi\left(\{y - (\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^\tau \tilde{\boldsymbol{\beta}} - \tilde{\bar{Y}}_k\}/\tilde{\sigma}\right). \tag{4}$$

By relaxing the equal area-specific error variance assumption, Jiang and Nguyen (2012) proposed an HNER model in which the area-specific variance of ε_{kj} is σ_k^2 and that of ν_k is $\gamma\sigma_k^2$. The corresponding EBLUP is given by

$$\breve{\bar{Y}}_k = \bar{\mathbf{X}}_k^\tau \breve{\boldsymbol{\beta}} + \frac{n_k \breve{\gamma}}{1 + n_k \breve{\gamma}} (\bar{y}_k - \bar{\mathbf{x}}_k^\tau \breve{\boldsymbol{\beta}}) \tag{5}$$

where $\breve{\boldsymbol{\beta}}$ and so on are MLEs under HNER. This leads to our second normal-model-based cdf estimator:

$$\breve{F}_k(y) \quad = \quad \frac{1}{n_k} \sum_{j=1}^{n_k} \Phi\left(\{y - (\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^\tau \breve{\boldsymbol{\beta}} - \breve{\bar{Y}}_k\}/\breve{\sigma}_k\right). \tag{6}$$

The corresponding small area quantile estimators will be referred to as the NER and HNER quantile estimators. Both are completely dependent on the normality assumption. Because of this, they were initially dismissed by the authors for the purpose of quantile estimation. Instead, the focus was on a semiparametric approach for estimating G_k, to be discussed in the next subsection. To our surprise, the performance of the NER- and HNER-based small area quantile estimations is satisfactory when the normality assumption holds, and it remains competitive when the model assumption is mildly violated.

3.2 Estimation under DRM

We now develop a third approach under a relaxed model assumption. We impose a DRM (Anderson 1979) on G_k, for $k = 1, 2, \dots, m$,

$$\log\{dG_k(t)/dG_0(t)\} = \boldsymbol{\theta}_k^\tau \mathbf{q}(t), \tag{7}$$

with a prespecified d_2-variate function $\mathbf{q}(t)$ and an area-specific tilting parameter $\boldsymbol{\theta}_k$. We require the first element of $\mathbf{q}(t)$ to be one, so the first element of $\boldsymbol{\theta}_k$ is a normalization parameter. The baseline distribution $G_0(t)$ is left unspecified, and $\mathbf{q}(t)$ could be chosen to be $(1,t)^\tau$. The nonparametric G_0 has flexibility, while the parametric tilting factor $\boldsymbol{\theta}_k^\tau \mathbf{q}(t)$ enables effective "strength borrowing" between small areas. Let us also emphasize that any G_j may be regarded as a baseline distribution because

$$\log\{dG_k(t)/dG_j(t)\} = (\boldsymbol{\theta}_k - \boldsymbol{\theta}_j)^\tau \mathbf{q}(t). \tag{8}$$

The only effect of the choice is to introduce a parameter transformation: $\boldsymbol{\theta}_k' = \boldsymbol{\theta}_k - \boldsymbol{\theta}_j$. The DRM is flexible, as is indicated by its inclusion of normal, Gamma, and other distribution families.

Unlike NER or HNER, the EL quantile estimates are linked to G_0, which will be made nonparametric here. An efficient nonparametric estimate of G_0, when available, results in efficient quantile estimates for all the small areas. At the same time, since it is nonparametric, with a proper choice of $q(t)$ this approach is likely robust to some degree of model mis-specification.

Empirical likelihood estimate of G_k

Consider an artificial situation where we have all the values of $\{\varepsilon_{kj} : j = 1, 2, \ldots, n_k\}$, $k = 0, \ldots, m$, from a DRM. These observations are the basis for inference on G_k. Following Owen (1988, 2001) or Qin and Lawless (1994), we confine the form of the candidate G_0 to $G_0(t) = \sum_{k,j} p_{kj} I(\varepsilon_{kj} \leqslant t)$, where $I(\cdot)$ is the indicator function and the summation $\sum_{k,j}$ is shorthand for $\sum_{k=0}^{m} \sum_{j=1}^{n_k}$. *The support of G_0 includes all ε_{kj}, not just those with $k = 0$.* This fact underlies the strength-borrowing strategy. In this setting, we have $p_{kj} = dG_0(\varepsilon_{kj})$ and $dG_k(\varepsilon_{ij}) = p_{ij} \exp\{\boldsymbol{\theta}_k^\tau \mathbf{q}(\varepsilon_{ij})\}$, $k = 0, 1, \ldots, m$, where the $\boldsymbol{\theta}_k$ are all d_2-variate unknown parameters. In other words, $G_k(t)$ is confined to the form

$$G_k(t) = \sum_{i,j} p_{ij} \exp\{\boldsymbol{\theta}_k^\tau \mathbf{q}(\varepsilon_{ij})\} I(\varepsilon_{ij} \leqslant t). \tag{9}$$

Clearly, $\boldsymbol{\theta}_0 = 0$ in the above expression. Because ε_{kj} follows $G_k(t)$, it contributes to the likelihood only through $dG_k(\varepsilon_{kj})$. The empirical likelihood (EL) is given by

$$L_n(G_0, G_1, \ldots, G_m) = \prod_{k,j} dG_k(\varepsilon_{kj}) = \big(\prod_{k,j} p_{kj}\big) \cdot \exp\big[\sum_{k,j}\{\boldsymbol{\theta}_k^\tau \mathbf{q}(\varepsilon_{kj})\}\big]$$

where the parameter $\boldsymbol{\theta}$ and the p_{kj}'s satisfy $p_{kj} \geqslant 0$ and for all $k = 0, 1, \ldots, m$,

$$\sum_{i,j} p_{ij} \exp\{\boldsymbol{\theta}_k^\tau \mathbf{q}(\varepsilon_{ij})\} = 1. \tag{10}$$

Note that the above summation is over i, j because k is reserved as the identity of the kth small area here. We will revert to k, j wherever possible.

Maximizing $\ell_n(\boldsymbol{\theta}, G_0)$ with respect to G_0 under the constraints (10) results in fitted probabilities (Qin and Lawless, 1994)

$$\hat{p}_{kj} = n^{-1}\{1 + \sum_{l=1}^{m} \lambda_l [\exp\{\boldsymbol{\theta}_l^\tau \mathbf{q}(\varepsilon_{kj})\} - 1]\}^{-1} \tag{11}$$

and the profile EL, up to an additive constant,

$$\tilde{\ell}_n(\boldsymbol{\theta}) = -\sum_{k,j} \log\{1 + \sum_{l=1}^{m} \lambda_l [\exp\{\boldsymbol{\theta}_l^\tau \mathbf{q}(\varepsilon_{kj})\} - 1]\} + \sum_{k,j}\{\boldsymbol{\theta}_k^\tau \mathbf{q}(\varepsilon_{kj})\}$$

with $(\lambda_1, \lambda_2, ..., \lambda_m)$ being the solution to

$$\sum_{i,j} \frac{\exp\{\boldsymbol{\theta}_k^\tau \mathbf{q}(\varepsilon_{ij})\} - 1}{1 + \sum_{l=1}^{m} \lambda_l [\exp\{\boldsymbol{\theta}_l^\tau \mathbf{q}(\varepsilon_{ij})\} - 1]} = 0$$

for $k = 1, \ldots, m$. The stationary point of $\tilde{\ell}_n(\boldsymbol{\theta})$ coincides with that of a dual form of the empirical log-likelihood function (Kezioua and Leoni-Aubina 2008):

$$\breve{\ell}_n(\boldsymbol{\theta}) = -\sum_{k,j} \log\big[\sum_{r=0}^{m} \rho_r \exp\{\boldsymbol{\theta}_r^\tau \mathbf{q}(\varepsilon_{kj})\}\big] + \sum_{k,j} \boldsymbol{\theta}_i^\tau \mathbf{q}(\varepsilon_{kj}), \tag{12}$$

with $\rho_r = n_r/n$, $r = 0, 1, \ldots, m$.

For point estimation, it is simpler to work with $\breve{\ell}_n(\boldsymbol{\theta})$, which is convex and free from constraints. Once the values of ε_{kj} are provided, we can easily find the maximum point, which serves as the maximum EL estimate of $\boldsymbol{\theta}$. It is then used to compute the fitted values defined by (11) with λ_l replaced by ρ_l. We subsequently obtain the estimator $\hat{G}_k$ and other parameters of interest via the invariance principle.

This approach first appears in Qin and Zhang (1997), Qin (1998), Zhang (1997), and others. In particular, the properties of the quantile estimators are discussed by Zhang (2000) and Chen and Liu (2013). For small area quantile estimation, however, we do not directly observe ε_{kj}. This difficulty is resolved by replacing these values by residuals obtained by fitting (1).

Parameter estimation for sampled small areas

Suppose we have independent samples representing $m+1$ small areas: $(y_{kj}, \mathbf{x}_{kj})$ for $k = 0, 1, \ldots, m$ and $j = 1, \ldots, n_k$. Under this model in which the distribution of y (or that of ϵ) is unspecified, we may minimize

$$\sum_{k,j}(y_{kj} - \nu_k - \mathbf{x}_{kj}^{\tau}\boldsymbol{\beta})^2$$

with respect to ν_k and $\boldsymbol{\beta}$ to obtain $\hat{\nu}_k = \bar{y}_k - \bar{\mathbf{x}}_k^{\tau}\hat{\boldsymbol{\beta}}$ with

$$\hat{\boldsymbol{\beta}} = \{\sum_{k,j}(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^{\tau}(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)\}^{-1}\{\sum_{k,j}(\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^{\tau}(y_{kj} - \bar{y}_k)\}, \tag{13}$$

where $\bar{\mathbf{x}}_k$ and $\bar{y}_k$ are sample means over small area k. The residuals of this fit are given by

$$\hat{\varepsilon}_{kj} = y_{kj} - \bar{y}_k - (\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^{\tau}\hat{\boldsymbol{\beta}}. \tag{14}$$

We then treat $\{\hat{\varepsilon}_{kj} : j = 1, 2, \ldots, n_k\}$ as samples from DRM and apply the EL method of Section 3.2.

Remark: Normality-based NER or HNER would have shrunk the predicted values of ν_k. This shrinkage will be postponed into the construction of $\hat{F}_k$ instead of occurring prematurely in $\hat{G}_k$.

Let $\ell_n(\boldsymbol{\theta})$ denote the log EL function (12) with ε_{kj} replaced by $\hat{\varepsilon}_{kj}$. We define the maximum EL estimator of $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}} = \text{argmax}\ell_n(\boldsymbol{\theta})$ and accordingly define the estimators

$$\hat{G}_k(t) = \sum_{i,j}\hat{p}_{ij}\exp\{\hat{\boldsymbol{\theta}}_k^{\tau}\mathbf{q}(\hat{\varepsilon}_{ij})\}I(\hat{\varepsilon}_{ij} < t) \tag{15}$$

with the convention $\hat{\boldsymbol{\theta}}_0 = \mathbf{0}$ and $\hat{p}_{ij} = n^{-1}\{1 + \sum_{l=1}^{m} \rho_l[\exp\{\hat{\boldsymbol{\theta}}_l^\tau \mathbf{q}(\hat{\varepsilon}_{ij})\} - 1]\}^{-1}$. This leads to EL-DRM-based cdf estimation, with some choice of $\hat{\bar{\mathbf{Y}}}$:

$$\hat{F}_k(y) = n_k^{-1} \sum_{j=1}^{n_k} \hat{G}_k(y - (\mathbf{x}_{kj} - \bar{\mathbf{x}}_k)^\tau \hat{\boldsymbol{\beta}} - \hat{\bar{\mathbf{Y}}}_k). \tag{16}$$

The corresponding small area quantile estimators will be referred to as EL estimators.

An authentic choice of $\hat{\bar{\mathbf{Y}}}_k$ is the regression estimator $\bar{y}_k + (\bar{\mathbf{X}}_k - \bar{\mathbf{x}}_k)\hat{\boldsymbol{\beta}}$. This choice amounts to utilizing $\hat{v}_k$ without shrinkage, and it was adopted in the first version of this paper. To better line up with the NER and HNER quantiles, we choose an NER-based $\tilde{\bar{\mathbf{Y}}}$ in the simulation. The difference between the two choices is negligible.

Because the basis of $\hat{G}_0$ (or equivalently any G_k) is on all n observations, the estimation is reliable. The estimation of the amount of tilting θ_k between them is largely done through direct observations. The low precision does not seem to cause serious damage in the estimation of F_k.

4 Properties of the EL quantile estimation

For each k, the covariates $\{\mathbf{x}_{kj}, j = 1, 2, \ldots, n_k\}$ are iid with a finite mean and a nonsingular and finite covariance matrix $\mathbf{V}_k$. The error terms $\{\varepsilon_{kj} : j = 1, 2, \cdots, n_k\}$ are iid samples, independent of the covariates, with conditional variance σ_k^2. The pure residuals ε_{kj} form $m + 1$ samples from populations with distribution function G_k satisfying (7). Let the total sample size $n = \sum_k n_k \to \infty$, and assume $\rho_k = n_k/n$ remains a constant (or within an n^{-1} range) as n increases. Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ be defined by (13).

Theorem 1. *Assume the general setting just presented in this section. Let* $\mathbf{V}_x = \sum_{k=0}^{m} \rho_k \mathbf{V}_k$. *As* $n \to \infty$, *we have* $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_\beta)$, *where* $\xrightarrow{d}$ *denotes convergence in distribution and* $\boldsymbol{\Sigma}_\beta = \mathbf{V}_x^{-1}(\sum_k \rho_k \mathbf{V}_k \sigma_k^2)\mathbf{V}_x^{-1}$.

For ease of exposition of the next theorem, we introduce some notation. For $k = 0, 1, \ldots, m$, let

$$h(x; \boldsymbol{\theta}) = \sum_{k=0}^{m} \rho_k \exp\{\boldsymbol{\theta}_k^{\tau} \mathbf{q}(x)\}; \quad h_k(x; \boldsymbol{\theta}) = \rho_k \exp\{\boldsymbol{\theta}_k^{\tau} \mathbf{q}(x)\}/h(x; \boldsymbol{\theta}).$$

Clearly, $0 < h_k < 1$ for all k. Let $\mathbf{h}(x; \boldsymbol{\theta}) = \{h_0(x; \boldsymbol{\theta}), \ldots, h_m(x; \boldsymbol{\theta})\}^{\tau}$ and define an $(m + 1) \times (m + 1)$ matrix

$$\mathbf{H}(x; \boldsymbol{\theta}) = \mathrm{diag}\{\mathbf{h}(x; \boldsymbol{\theta})\} - \mathbf{h}(x; \boldsymbol{\theta})\mathbf{h}^{\tau}(x; \boldsymbol{\theta}).$$

We will use $h(x; \dot{x})$ and $h(x; \dot{\boldsymbol{\theta}})$ for the partial derivatives of h with respect to x and $\boldsymbol{\theta}$ respectively. When $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, the true value of $\boldsymbol{\theta}$, we may drop $\boldsymbol{\theta}^*$ in $h(x; \boldsymbol{\theta}^*)$ and denote it as $h(x)$. Lastly, we use $d\bar{G}(x)$ for $h(x; \boldsymbol{\theta}^*) dG_0(x)$ in the integrations.

Theorem 2. *Assume the conditions of Theorem 1. Furthermore, the population distributions G_k satisfy DRM (7) with true parameter value $\boldsymbol{\theta}^*$, and $\int h(t; \boldsymbol{\theta}) dG_0 < \infty$ in a neighborhood of $\boldsymbol{\theta}^*$. Also assume that the components of $\mathbf{q}(t)$ are linearly independent with the first element being one, they are twice differentiable, and there exist a function $\psi(t)$ and $c_0 > 0$ such that*

$$\sup_{t:|t-u| \leqslant c_0} \{\|\mathbf{q}(t)\| \|\mathbf{q}(\ddot{t})\| + \|\mathbf{q}(t)\| \|\mathbf{q}(\dot{t})\|^2\} \leqslant \psi(u) \tag{17}$$

for all u with $\int \psi(u) d\bar{G}(u) < \infty$. Then as n goes to infinity, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega}$ is given in (A.10) in the supplementary material.

The assumption that $\int h(t; \boldsymbol{\theta}) dG_0(t) < \infty$ in a neighborhood of $\boldsymbol{\theta}^*$ implies the existence of the moment generating function of $\mathbf{q}(t)$ and therefore all its finite moments. Note that our log EL is

$$\ell_n(\boldsymbol{\theta}) = \sum_{k,j} [\boldsymbol{\theta}_k^{\tau} \mathbf{q}(\hat{\varepsilon}_{kj}) - \log\{h(\hat{\varepsilon}_{kj}; \boldsymbol{\theta})\}]. \tag{18}$$

Since the first component of $\mathbf{q}(t)$ is equal to one, the condition in (17) implies

$$\sup_{t:|t-u| \leqslant c_0} \|\mathbf{q}(\ddot{t})\| + \|\mathbf{q}(\dot{t})\|^2 \leqslant \psi(u) \text{ for all } u. \tag{19}$$

Therefore, for all $\{k, j\}$, when we approximate $\hat{\boldsymbol{\theta}}_k^{\tau}\mathbf{q}(\hat{\varepsilon}_{kj}) - \log\{h(\hat{\varepsilon}_{kj}; \hat{\boldsymbol{\theta}})\}$ by their first-order Taylor expansions at $(\varepsilon_{kj}; \boldsymbol{\theta}^*)$, the summation of the remainders is of the order $o_p(1)$. Combining this with the proof of Theorem 2.1 of Chen and Liu (2013) immediately leads to the root-n consistency of $\hat{\boldsymbol{\theta}}$. This observation alleviates the burden of proof.

We next examine the asymptotic properties of the proposed small area quantile estimators, which we call EL quantiles for short.

Theorem 3. *Assume the conditions of Theorem 2. Suppose in addition that the $G_k(t)$ have smooth and bounded density functions, and $F_k(y)$ has positive density at ξ_k. Then the EL quantile in (2) is root-n consistent. That is, $\hat{\xi}_k - \xi_k = O_p(n^{-1/2})$.*

5 Simulation study

In this section, we use simulation to demonstrate the performance of the small area quantile estimators. Ideally, we should compare the new methods with all existing methods including those of Chambers and Tzavidis (2006), Molina and Rao (2010), and Chaudhuri and Ghosh (2011). As pointed out earlier, the M-quantile method is possible only if all the covariates are observed. All of them are computationally involved, which introduces substantial obstacles in their independent implementation. Because of this, we limit our simulation to four quantile estimators. One is the "Direct" estimator based on the empirical quantiles. The other three are the quantiles of (4) and (6) (NER and HNER for short) and the EL quantiles. In the simulation, we examine the 5%, 25%, 50%, 75%, and 95% small area quantile estimations.

5.1 Simulation settings

The first task of the simulation is to create finite populations. We consider the following two models for the general structure of the population:

$$y_{kj} = \mathbf{x}_{kj}^{\tau}\boldsymbol{\beta} + \nu_k + \varepsilon_{kj}, \tag{A}$$

$$y_{kj} = \mathbf{x}_{kj}^{\tau}\boldsymbol{\beta} + 0.4\log(1 + \mathbf{x}_{kj}^{\tau}\mathbf{x}_{kj}/100) + \nu_k + \varepsilon_{kj}. \tag{B}$$

Model (A) contains a linear structure between the response variable y and the predetermined covariate vector $\mathbf{x}$. Model (B) contains in addition a nonlinear structure that helps us to examine the model robustness of the quantile estimators.

To ensure a high level of authenticity, we use real survey data (to be presented in the next subsection) as the blueprint for other details of the finite populations:

1. sample $N_k = 1000$ three-dimensional $\mathbf{x}_{kj}$ values without replacement from the real survey data, $k = 0, 1, \ldots 19$;
2. let $\boldsymbol{\beta}_0^{\tau} = (0.021, 0.024, 0.070)$, the fitted value from the real survey data;
3. generate ν_k from $N(8, 1)$ and from $N(8, 4)$ to mimic the real-data fit.

Next, we specify the error distribution, that of ε_{kj}. We sampled ε_{kj} from

(i) $N(0, 1)$;

(ii) normal mixture $0.5N(-\mu_k/6, 1) + 0.5N(\mu_k/6, 1)$;

(iii) normal mixture $0.1N(-\mu_k/2, 1) + 0.9N(\mu_k/18, 1)$;

(iv) normal mixture $0.9N(-\mu_k/18, 1) + 0.1N(\mu_k/2, 1)$;

(v) fitted residuals based on the real survey data after inflation by a factor of 1.15.

Note that only one of the above five error distributions is employed for each finite population created. Choice (i) matches the NER model assumption; (ii) is non-normal but symmetric; (iii) and (iv) are skewed in opposite directions; and (v) is from real data. The shape of the distribution in (v) is close to that of the skewed distribution in (iii).

To induce heterogeneity while avoiding too much human intervention, we generated μ_k in (ii)–(iv) from a uniform distribution on the interval $[4.5, 6]$. Another factor is the signal-to-noise ratio, i.e., the proportion of the within-small-area variation explained by the linear structure $\mathbf{x}\boldsymbol{\beta}$. We generated finite populations with $\boldsymbol{\beta} = 1.0\boldsymbol{\beta}_0, 1.25\boldsymbol{\beta}_0$, and $1.5\boldsymbol{\beta}_0$. These choices set the signal-to-noise ratios to around 30%, 50%, and 70%. Some choices may still appear unnatural due to our desire to enforce similar σ^2 values and so on across the populations or subpopulations after the other parameter values have been chosen. Once a finite population is generated/created, the small area population quantiles are computed and regarded as target parameters for the inference.

The combination of 2 model structures (A and B), 5 types of error distribution, 3 scales in $\boldsymbol{\beta}$, and 2 choices for the v_k-generating distribution leads to 60 finite populations. These are further compounded with 2 sample sizes, $n = 200$ and $n = 1000$. From these results, we will present six sets by omitting the results with $\boldsymbol{\beta} = 1.5\boldsymbol{\beta}_0$ and $v_k \sim N(8, 4)$ and so on.

With $N = 1000$ repetitions, we drew random samples of size n without replacement from the finite population and estimated the parameters. We did not completely follow simple random sampling without replacement at the population level. To avoid the possibility that some small areas have very few sampled units, we drew $n - 40$ units randomly at the population level and allocated an additional 2 units in each small area. We need $n_k \geqslant 2$ to properly estimate $\boldsymbol{\theta}_k$. This plan is not convenient in a real-world situation. However, it does not distort our comparison of the methods and prevents ad hoc adjustments. In real-world applications, one may have to set small areas with $n_k < 2$ aside and treat them separately.

We experimented with many choices of $\mathbf{q}(t)$ in the DRM and eventually settled on $\mathbf{q}_1(t) = (1, t)^\tau$ and

$\mathbf{q}_2(t) = (1, \text{sign-root}(t))^\tau$. The corresponding EL quantiles will be referred to as EL(1) and EL(2). These two basis functions are not "true." The choices are motivated by the general overall performance in terms of "model robustness." Let $\hat{\xi}_k^{(j)}$ be a generic quantile estimate in the jth repetition. We report the average mean squared error (AMSE), defined to be

$$\text{AMSE} = \{N(m+1)\}^{-1} \sum_{k=0}^{m} \sum_{j=1}^{N} (\hat{\xi}_k^{(j)} - \xi_k)^2.$$

When the model does not match the population, bias can inflate the size of AMSE. For this reason, we also report the average absolute biases:

$$\text{ABIAS} = (m+1)^{-1} \sum_{k=0}^{m} \left| N^{-1} \sum_{j=1}^{N} \hat{\xi}_k^{(j)} - \xi_k \right|.$$

Because it fixes the finite population and average over repeated implementations of the sampling plan, the AMSE is a criterion for design-based analysis. Many of our theoretical results are for model-based analysis. Based on pilot simulation trials, the AMSE comparisons do not change from one basis to another. Design-based simulation avoids the substantial computational burden of repeatedly generating large finite populations and computing their parameter values. This harmless compromise saves a great deal of computation.

5.2 Simulation results

Tables 1–6 present the simulated AMSE and ABIAS values under Models (A) and (B) with the error distributions (i)–(v) for the five quantile estimators. Recall that EL(1) and EL(2) are EL quantiles with two choices of the basis function: $\mathbf{q}_1(t) = (1, t)^\tau$ and $\mathbf{q}_2(t) = (1, \text{sign-root}(t))^\tau$.

Across all six tables and the omitted results, we observe that our three methods have an AMSE that is substantially lower than that of the Direct method. This is an indication that the strategy of "strength-borrowing" works. When the error distribution is homogenous normal (i), NER generally has a noticeable lower AMSE. The superiority of NER under ("incorrect") model (B) is still noticeable but more moderate; see Tables 4–6. EL is a close second under error distribution (i).

Let us now focus on the performance under error distribution (ii): non-normal but symmetric with mild heterogeneity. It is clear that both Direct and HNER lose out. NER and EL(2) now take turns at having the lowest AMSE. EL(1) is in a respectable third place. HNER remains the last, but its performance is close to that of the other methods in many instances. The nonlinear structure of (B) inflates the AMSE of all the methods.

Next, we move to the results for the populations with other error distributions: they are all skewed. NER remains largely competitive for 50% quantile estimation and occasionally has good performance for 5% quantile estimation (for instance, Model A (iv, v)). The two EL quantiles are routinely the best. For Model (A) error distribution (iii), the AMSE values of the ELs are 0.307 and 0.278, compared with 0.708 and 1.326 for NER and HNER. HNER generally has the highest AMSE, although there are a few exceptions (Model A (iv) for the 25% quantile). Of the two EL quantiles, EL(1) works better in most cases under the skewed error distributions (iii)–(v).

Table 2 contains results parallel to those in Table 1, where the population regression coefficient vector is inflated by a factor of 1.25. This change increases the signal of the linear structure and improves the precision of all the statistical methods. However, it also increases the scale of y and therefore those of the subpopulation quantiles. We cannot easily quantify the effect of this change. However, we find that it has no effect on the relative performance of the five quantiles.

Increasing the sample size in general reduces both the bias and the variance of point estimators. Tables 1 and 3 are the same except for different sample sizes: $n = 200$ and $n = 1000$. In Table 3 we see a substantially reduced AMSE in all the methods. The level of reduction is, however, not uniform. Direct and the ELs are clear winners under error distributions (iii)–(v). This is because NER and HNER are constructed based on normality, so they are likely inconsistent in an asymptotic sense. Their biases do not become zero even if the sample size increases to infinity. In the simulations, the ABIAS values of NER and HNER do not decrease as fast as those of the EL methods.

Model (B) differs from Model (A) by having a nonlinear structure. This structure “invalidates” all the methods except the Direct method. Tables 4, 5, and 6 give the results for model B corresponding to Tables 1, 2, and 3. The nonlinear population structure hurts EL(1), EL(2), NER, and HNER indiscriminately. However, the discussions from Model (A) remain applicable.

Bias can be a major factor in the AMSE. The superior performance of the ELs compared with NER or HNER under error distributions (iii)–(v) is attributable to their lower biases (except for some 5% quantiles). Model (B) contains an additive nonlinear structure compared to Model (A). The EL methods are semiparametric, and we expected that this flexibility would make them stronger competitors under mild structural mis-specification. Our simulation results do not seem to support this conjecture.

A more streamlined summary is as follows.

1. When the error distribution is normal or non-normal but symmetric, EL(1), EL(2), and NER have the lowest AMSE. NER is marginally the best.

2. If the error distribution is skewed to the left (right), then the 5% (95%) quantile is the hardest to estimate with good precision. Excluding this quantile, the EL quantiles have a substantially lower AMSE.

3. Suppose the results in Table 1 are typical. The reduction of AMSE from NER to EL(1) is 44% on average over (iii)–(v) after excluding the nonperforming 5% quantiles of (iii) and (v) and the 95% quantile of (iv). That is, EL(1) outperforms by a large margin.

4. The nonlinear population structure prescribed by Model (B) hurts EL(1), EL(2), NER, and HNER indiscriminately. It does not alter their relative performance.

5. Increasing the importance of the linear structure, from $\boldsymbol{\beta}_0$ to $1.25\boldsymbol{\beta}_0$ or $1.5\boldsymbol{\beta}_0$ (this case omitted), does not alter the relative performance of these quantile estimators.

6. The EL methods benefited the most from an increased sample size.

7. When the error distribution is skewed to the left as in (iii) and (v), the AMSE for estimating 5% is very large. When it is skewed to the right as in (iv), the 95% quantile is the hardest to estimate. These results can be explained by the lower value of the density functions.

8. Bias plays a larger role in NER and HNER in terms of the AMSE when the sample size increases. This points to the model dependence of these two methods.

A straightforward summary is: NER works best and the EL methods are comparable when the error distribution is normal or nearly normal. The EL methods outperform significantly when the error distribution is skewed.

5.3 Simulation based on real data

Finite populations created based on statistical models are inevitably artificial. We wish to study our method on a population that is closer to the real world. For this reason, we downloaded the data set *Survey of Labour and Income Dynamics* (SLID) provided by Statistics Canada (2014) from the University of British Columbia library data centre. According to the readme file, this survey complements traditional survey data on labour market activity and income with an additional dimension: the changes experienced by individuals over time.

We are grateful to Statistics Canada for making the data set available, but we do not address the original goal of the survey here. Instead, we use it to create a realistic finite population to study the effectiveness of our small area quantile estimator.

The data set contains 147 variables and 47705 sampling units. We keep 6 of the 147 variables, i.e., total income (`ttin`), gender (`gender`), age (`age`), years of experience (`yrx`), number of weeks employed (`tweek`), and education level (`edu`). More precise definitions are not necessary here. We remove any units containing missing values in these six variables as well as those with $\texttt{ttin} \leqslant 0$. The problem of estimating the proportion of zeroes in a population is a research topic by itself and is not the focus of this paper. Their removal makes the population easy to handle without compromising the authenticity that we wish to establish. The resulting data set contains 35488 sampling units.

We ignore the sampling plan under which this data set was obtained. Instead, we examine how well our proposed small area quantile estimators perform if we sample from this "real" population. We first create 10 age groups composed of individuals whose ages are in the following intervals:

[0, 20)	[20, 25)	[25, 30)	[30, 35)	[35, 40)	[40, 45)	[45, 50)	[50, 55)	[55, 60)	[60, ∞)

Each age group is then divided into male and female subpopulations. This forms a finite population with 20 small domains (i.e., small areas) based on age–gender combinations. The sizes of these small domains are as follows.

Male	1231	1525	1372	1337	1469	1536	1866	1890	1920	3089
Female	1200	1433	1449	1504	1497	1695	2053	2019	1944	3459

We regard $\log(1 + \texttt{ttin})$ as the response variable and use `yrx`, `tweek`, and `edu` as covariates. The small area 5%, 25%, 50%, 75%, and 95% population quantiles are depicted in Figure 1 for males and females separately. We use simulation to evaluate how well they are estimated by our three methods. We report the results for $n = 200$ and $n = 1000$ in Table 7.

We first notice that the 5% quantile is the hardest to estimate with good precision. This is because the error distribution is skewed toward the left. This density function has lower densities at lower quantiles. The variance of a quantile estimator is generally inversely proportional to the size of the density function.

Hence, all the methods have a large AMSE compared with the AMSE at other quantiles. Overall, HNER is the worst if Direct is not counted. None of the other methods stand out.

For the other four quantiles, EL(1) has the lowest AMSE and EL(2) is a close second. The improvement of the EL methods over NER or HNER is substantial. For the median estimation, the AMSE values of NER are 40% ($n = 200$) and 180% ($n = 1000$) higher than those for EL(1). The comparisons for the other quantiles (excluding 5%) are just as sharp.

Because the data set is not generated from a model with a linear structure, we believe that increasing the sample size should reduce the variance of all the estimators significantly but have a smaller effect on the biases. However, the biases of the EL quantiles should be reduced to a larger degree because of their flexible nonparametric error distribution assumption. This is the case for the 25%, 50%, and 75% quantile estimations. We observe a 65% reduction in bias for EL(1) compared with merely 20% for NER. The changes in the biases of the 5% and 95% quantile estimations are small.

In Figure 2, we depict the the small area median estimates of EL(1) and NER when $n = 200$. The y-axis has been transformed to `ttin`, the total income (presumably in Canadian dollar). For each small area, we draw a box where the top, middle, and bottom lines are 90th, 50th, and 10th percentiles of 1000 median estimates of EL(1) and NER respectively. The true values of the small area (age–gender group) medians are marked by stars. We find that all the EL(1) boxes contain the true small area medians, but two NER boxes fail for two small areas. The EL(1) boxes are generally shorter, with the stars located closer to the middle line.

Based on these results, we conclude that the EL quantiles outperform NER and HNER by a large margin for this population where the error distribution is skewed (based on $\log(1 + \texttt{ttin})$).

6 Conclusions and discussion

We have developed three methods for small area quantile estimation. All of them are based on a linear regression model at the unit level. The first two are based on a normal error distribution, and the third is based on a semiparametric DRM via empirical likelihood.

All three methods are helpful compared to a direct estimator constructed through the area-specific empirical distributions. In particular, the EL quantile estimators are semiparametric, and their performance is satisfactory over a broad range of error distributions. When the error distribution is skewed, the EL quantiles are by far the best. The NER quantile estimator is surprisingly resistant to mild violation of the normality assumption. The performance of the HNER quantile estimator is less satisfactory in the examples we investigated.

In theory, the EL quantiles are consistent only when the observations are independent. Independence is approximately satisfied when the sample fraction is low. There are no practical obstacles to the application of the new methods to data collected via complex sampling plans, although we need to work on a solid justification.

In statistical practice, it is generally desirable to be able to assess the precision of an estimator through a good MSE estimation. For the EL quantile this is likely to be a challenging task. We plan to discuss the MSE issue in a future paper.

This paper raises more questions than it has answered. In particular, the theory totally ignores randomness due to the probability sampling plan or the population structure. In addition, the EL approach should have a built-in smoothing mechanism on the estimation of $\boldsymbol{\theta}_k$ across the small areas. The asymptotic results based on a fixed number of small areas are merely placebos. All these shortcomings are serious but should be overcome by our continued efforts.

Supplemantary material

The supplementary material presents the proofs of Theorems 1–3.

References

Anderson, J. A. (1979). Multivariate logistic compounds. *Biometrika* 66: 17–26.

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* 80: 28–36.

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

Chambers, R. L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika* 73: 597–604.

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

Chaudhuri, S. and Ghosh, M. (2011). Empirical likelihood for small area estimation. *Biometrika* 98: 473–480.

Chen, J. and Liu, Y. (2013). Quantile and quantile-function estimations under density ratio model. *The Annals of Statistics* 4: 1669–1692.

Chen, J., Rao, J. N. K. and Sitter, R.R. (2000). Efficient random imputation for missing data in complex surveys. *Statistica Sinica* 10: 1153–1169.

Chen, J. and Wu, C. (2002). Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. *Statistica Sinica* 12: 1223–1239.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74: 269–277.

Francisco, C. A. and Fuller, W. A. (1991). Quantile estimation with a complex survey design. *The Annals of Statistics* 19: 454–469.

Ghosh, M., Maiti, T. and Roy, A. (2008). Influence functions and robust Bayes and empirical Bayes small area estimation. *Biometrika* 95: 573–585.

Jiang, J. and Lahiri, P. S. (2006). Estimation of finite population domain means: A model-assisted empirical best prediction approach. *Journal of the American Statistical Association* 101: 301–311.

Jiang, J. and Nguyen, T. (2012). Small area estimation via heteroscedastic nested-error regression. *The Canadian Journal of Statistics* 40: 588–603.

Jiang, J., Nguyen, T. and Rao, J. S. (2010). Fence method for nonparametric small area estimation. *Survey Methodology* 36: 3–11.

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

Kezioua, A. and Leoni-Aubina, S. (2008). On empirical likelihood for semiparametric two-sample density ratio models. *Journal of Statistical Planning and Inference* 138: 915–928.

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

Kriegler, B. and Berk, R. (2010). Small area estimation of the homeless in Los Angeles: An application of cost-sensitive stochastic gradient boosting. *The Annals of Applied Statistics* 4: 1234–1255.

Lahiri, P. S. and Rao, J. N. K. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association* 90: 758–766.

Marchetti, S., Tzavidis, N. and Pratesi, M. (2012). Non-parametric bootstrap mean squared error estimation for m-quantile estimators of small area averages, quantiles and poverty indicators. *Computational Statistics and Data Analysis* 56: 2889–2902.

Molina, I. and Rao, J. N. K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics* 38: 369–385.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75: 237–249.

Owen, A. B. (2001). *Empirical Likelihood*. New York: Chapman and Hall/CRC.

Pfeffermann, D. (2002). Small area estimation-New developments and directions. *International Statistical Review* 70: 125–143.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science* 28: 40–68.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association* 85: 163–171.

Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika* 85: 619–630.

Qin, J. and Lawless, J. F. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics* 22: 300–325.

Qin, J. and Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika* 84: 609–618.

Rao, J. N. K. (2003). *Small Area Estimation*. New York: Wiley.

Schaible, W. L. (1993). Use of small area estimators in U.S. federal programs. In *Small Area Statistics and Survey Designs* edited by G. Kalton, J. Kordos and R. Platek, 1: 95–114. Central Statistical Office, Warsaw.

Statistics Canada (2014). *Survey of Labour and Income Dynamics, 2011*. Access: ABACUS. `http://hdl.handle.net/10573/42961`.

Tzavidis, N. and Chambers, R. (2005). Bias adjusted estimation for small areas with M-quantile models. *Statistics in Transition* 7: 707–713.

Tzavidis, N., Salvati, N., Monia Pratesi, M. and Chambers, R. (2007). M-Quantile Models with Application to Poverty Mapping. The University of Manchester. *CCSR Working Paper* 2007-05.

Wang, S. and Dorfman, A. H. (1996). A new estimator for the finite population distribution function. *Biometrika* 83: 639–652.

Zhang, B. (1997). Assessing goodness-of-fit of generalized logit models based on case-control data. *Journal of Multivariate Analysis* 82: 17–38.

Zhang, B. (2000). Quantile estimation under a two-sample semiparametric model. *Bernoulli* 6: 491–511.

Table 1: AMSE and ABIAS of small area quantile estimators (Model A, $n = 200$, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$).

		AMSE					ABIAS				
Error	α	Direct	EL(1)	EL(2)	NER	HNER	Direct	EL(1)	EL(2)	NER	HNER
	5%	0.588	0.157	0.152	0.133	0.264	0.173	0.150	0.147	0.111	0.176
	25%	0.264	0.109	0.103	0.100	0.129	0.086	0.085	0.083	0.073	0.099
(i)	50%	0.195	0.107	0.100	0.099	0.100	0.022	0.082	0.080	0.079	0.080
	75%	0.220	0.115	0.109	0.106	0.119	0.071	0.092	0.092	0.087	0.077
	95%	0.455	0.145	0.141	0.126	0.231	0.149	0.142	0.142	0.119	0.111
	5%	0.436	0.192	0.166	0.147	0.216	0.169	0.144	0.142	0.127	0.143
	25%	0.236	0.156	0.117	0.124	0.136	0.054	0.103	0.101	0.128	0.139
(ii)	50%	0.302	0.182	0.120	0.117	0.118	0.061	0.086	0.085	0.090	0.093
	75%	0.238	0.175	0.130	0.156	0.168	0.032	0.139	0.134	0.198	0.211
	95%	0.243	0.150	0.130	0.126	0.201	0.141	0.097	0.095	0.092	0.100
	5%	1.662	0.307	0.278	0.708	1.326	0.478	0.351	0.304	0.733	0.817
	25%	0.603	0.110	0.136	0.243	0.372	0.258	0.092	0.090	0.341	0.299
(iii)	50%	0.135	0.093	0.116	0.174	0.179	0.036	0.102	0.100	0.248	0.239
	75%	0.084	0.099	0.118	0.111	0.060	0.028	0.090	0.089	0.078	0.090
	95%	0.144	0.117	0.132	0.336	0.291	0.090	0.104	0.107	0.448	0.335
	5%	0.339	0.163	0.179	0.249	0.229	0.150	0.113	0.113	0.322	0.233
	25%	0.152	0.096	0.116	0.123	0.066	0.082	0.084	0.085	0.126	0.092
(iv)	50%	0.112	0.092	0.115	0.138	0.134	0.016	0.082	0.082	0.157	0.166
	75%	0.518	0.105	0.131	0.275	0.390	0.236	0.097	0.098	0.384	0.346
	95%	1.980	0.437	0.354	1.032	1.840	0.683	0.451	0.379	0.918	1.038
	5%	1.762	0.283	0.288	0.258	0.672	0.125	0.263	0.258	0.240	0.249
	25%	0.360	0.100	0.119	0.165	0.264	0.151	0.115	0.117	0.224	0.164
(v)	50%	0.144	0.087	0.105	0.125	0.127	0.027	0.104	0.104	0.157	0.146
	75%	0.132	0.098	0.114	0.114	0.086	0.052	0.100	0.100	0.108	0.096
	95%	0.293	0.142	0.156	0.272	0.279	0.078	0.166	0.165	0.347	0.192

Table 2: AMSE and ABIAS of small area quantile estimators (Model A, $n = 200$, $\boldsymbol{\beta} = 1.25\boldsymbol{\beta}_0$).

		AMSE					ABIAS				
Error	α	Direct	EL(1)	EL(2)	NER	HNER	Direct	EL(1)	EL(2)	NER	HNER
	5%	0.703	0.190	0.185	0.165	0.285	0.177	0.158	0.156	0.127	0.181
	25%	0.314	0.113	0.107	0.104	0.132	0.099	0.088	0.086	0.078	0.098
(i)	50%	0.220	0.108	0.101	0.100	0.102	0.027	0.080	0.077	0.075	0.077
	75%	0.239	0.119	0.113	0.110	0.121	0.073	0.089	0.089	0.085	0.075
	95%	0.474	0.152	0.148	0.134	0.233	0.146	0.140	0.140	0.117	0.105
	5%	0.591	0.238	0.209	0.177	0.240	0.194	0.156	0.155	0.138	0.156
	25%	0.285	0.154	0.115	0.120	0.132	0.089	0.087	0.086	0.098	0.108
(ii)	50%	0.316	0.179	0.120	0.121	0.122	0.065	0.092	0.092	0.102	0.105
	75%	0.270	0.182	0.135	0.159	0.170	0.044	0.137	0.132	0.192	0.203
	95%	0.271	0.162	0.141	0.132	0.204	0.150	0.103	0.101	0.081	0.098
	5%	1.574	0.266	0.254	0.542	1.069	0.363	0.300	0.267	0.586	0.655
	25%	0.650	0.129	0.156	0.217	0.347	0.252	0.095	0.094	0.283	0.246
(iii)	50%	0.178	0.098	0.121	0.183	0.191	0.056	0.106	0.105	0.260	0.247
	75%	0.100	0.108	0.127	0.116	0.068	0.028	0.094	0.093	0.082	0.095
	95%	0.158	0.129	0.144	0.325	0.283	0.095	0.106	0.108	0.426	0.323
	5%	0.482	0.224	0.240	0.226	0.217	0.188	0.134	0.134	0.231	0.165
	25%	0.222	0.108	0.128	0.135	0.080	0.111	0.098	0.099	0.145	0.115
(iv)	50%	0.138	0.096	0.119	0.131	0.125	0.024	0.076	0.077	0.129	0.141
	75%	0.498	0.112	0.139	0.258	0.367	0.219	0.096	0.097	0.352	0.320
	95%	1.979	0.456	0.375	0.982	1.753	0.662	0.463	0.389	0.873	0.990
	5%	1.771	0.285	0.293	0.268	0.652	0.097	0.251	0.246	0.221	0.230
	25%	0.425	0.112	0.131	0.166	0.263	0.165	0.125	0.126	0.211	0.158
(v)	50%	0.175	0.092	0.109	0.130	0.135	0.035	0.108	0.109	0.168	0.153
	75%	0.149	0.105	0.121	0.117	0.091	0.049	0.099	0.098	0.101	0.096
	95%	0.309	0.155	0.169	0.271	0.277	0.091	0.168	0.168	0.334	0.180

Table 3: AMSE and ABIAS of small area quantile estimators (Model A, $n = 1000$, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$).

		AMSE					ABIAS				
Error	α	Direct	EL(1)	EL(2)	NER	HNER	Direct	EL(1)	EL(2)	NER	HNER
	5%	0.141	0.033	0.032	0.031	0.048	0.087	0.067	0.065	0.063	0.060
	25%	0.048	0.022	0.021	0.021	0.025	0.015	0.035	0.033	0.031	0.034
(i)	50%	0.038	0.022	0.020	0.020	0.020	0.010	0.027	0.026	0.027	0.027
	75%	0.042	0.023	0.022	0.021	0.024	0.016	0.039	0.039	0.039	0.029
	95%	0.105	0.030	0.029	0.028	0.045	0.081	0.067	0.067	0.065	0.041
	5%	0.104	0.034	0.029	0.027	0.035	0.080	0.048	0.047	0.058	0.059
	25%	0.039	0.024	0.018	0.023	0.025	0.014	0.028	0.028	0.070	0.070
(ii)	50%	0.051	0.030	0.019	0.022	0.022	0.016	0.029	0.028	0.061	0.062
	75%	0.044	0.027	0.019	0.037	0.039	0.015	0.032	0.033	0.132	0.132
	95%	0.040	0.025	0.022	0.024	0.033	0.043	0.034	0.033	0.065	0.066
	5%	0.292	0.032	0.033	0.273	0.346	0.094	0.082	0.062	0.485	0.493
	25%	0.072	0.017	0.023	0.084	0.104	0.038	0.033	0.025	0.249	0.244
(iii)	50%	0.022	0.013	0.017	0.060	0.061	0.009	0.026	0.027	0.204	0.203
	75%	0.016	0.015	0.019	0.017	0.009	0.007	0.028	0.030	0.027	0.022
	95%	0.031	0.018	0.021	0.173	0.168	0.039	0.037	0.038	0.389	0.373
	5%	0.083	0.033	0.037	0.089	0.084	0.060	0.055	0.055	0.246	0.238
	25%	0.027	0.015	0.019	0.025	0.015	0.016	0.025	0.026	0.077	0.072
(iv)	50%	0.019	0.014	0.018	0.035	0.034	0.006	0.019	0.018	0.128	0.130
	75%	0.034	0.015	0.020	0.112	0.131	0.024	0.025	0.027	0.301	0.297
	95%	0.335	0.044	0.043	0.539	0.636	0.067	0.129	0.110	0.695	0.709
	5%	0.580	0.123	0.120	0.118	0.178	0.253	0.240	0.225	0.208	0.160
	25%	0.058	0.020	0.026	0.075	0.098	0.027	0.047	0.055	0.216	0.198
(v)	50%	0.024	0.015	0.020	0.041	0.042	0.009	0.027	0.031	0.141	0.136
	75%	0.023	0.022	0.026	0.030	0.018	0.011	0.055	0.058	0.088	0.044
	95%	0.079	0.052	0.056	0.181	0.157	0.083	0.152	0.151	0.362	0.300

Table 4: AMSE and ABIAS of small area quantile estimators (Model B, $n = 200$, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$).

		AMSE					ABIAS				
Error	α	Direct	EL(1)	EL(2)	NER	HNER	Direct	EL(1)	EL(2)	NER	HNER
(i)	5%	1.003	0.299	0.293	0.265	0.380	0.204	0.220	0.218	0.185	0.226
	25%	0.441	0.132	0.126	0.122	0.152	0.140	0.097	0.095	0.097	0.107
	50%	0.263	0.114	0.107	0.106	0.110	0.026	0.080	0.079	0.078	0.079
	75%	0.257	0.130	0.124	0.120	0.130	0.072	0.090	0.088	0.084	0.076
	95%	0.487	0.171	0.167	0.154	0.251	0.152	0.132	0.133	0.115	0.115
(ii)	5%	0.966	0.380	0.347	0.285	0.345	0.262	0.229	0.229	0.197	0.199
	25%	0.397	0.160	0.123	0.131	0.143	0.135	0.077	0.078	0.085	0.082
	50%	0.354	0.176	0.121	0.125	0.128	0.065	0.088	0.088	0.103	0.107
	75%	0.302	0.200	0.151	0.177	0.189	0.051	0.159	0.153	0.210	0.221
	95%	0.282	0.184	0.163	0.152	0.225	0.153	0.107	0.106	0.095	0.107
(iii)	5%	1.608	0.237	0.236	0.415	0.843	0.209	0.223	0.209	0.381	0.432
	25%	0.816	0.209	0.237	0.234	0.372	0.233	0.184	0.182	0.271	0.253
	50%	0.256	0.121	0.143	0.214	0.228	0.081	0.148	0.146	0.305	0.288
	75%	0.111	0.126	0.146	0.129	0.087	0.023	0.111	0.108	0.099	0.118
	95%	0.160	0.152	0.167	0.343	0.304	0.098	0.113	0.115	0.432	0.331
(iv)	5%	0.882	0.421	0.436	0.289	0.303	0.279	0.307	0.305	0.157	0.187
	25%	0.401	0.157	0.179	0.189	0.137	0.176	0.158	0.159	0.234	0.198
	50%	0.174	0.106	0.130	0.123	0.115	0.032	0.072	0.073	0.099	0.103
	75%	0.476	0.125	0.150	0.253	0.353	0.200	0.088	0.087	0.330	0.302
	95%	2.047	0.493	0.416	0.966	1.715	0.672	0.488	0.420	0.853	0.960
(v)	5%	1.933	0.403	0.412	0.389	0.757	0.151	0.318	0.316	0.277	0.322
	25%	0.590	0.158	0.176	0.196	0.288	0.194	0.159	0.158	0.242	0.179
	50%	0.230	0.104	0.122	0.147	0.155	0.047	0.114	0.115	0.192	0.173
	75%	0.164	0.123	0.139	0.131	0.106	0.045	0.109	0.109	0.114	0.107
	95%	0.319	0.187	0.201	0.310	0.298	0.093	0.190	0.190	0.370	0.197

Table 5: AMSE and ABIAS of small area quantile estimators (Model B, $n = 200$, $\boldsymbol{\beta} = 1.25\boldsymbol{\beta}_0$).

		AMSE					ABIAS				
Error	α	Direct	EL(1)	EL(2)	NER	HNER	Direct	EL(1)	EL(2)	NER	HNER
(i)	5%	1.229	0.406	0.399	0.364	0.472	0.254	0.267	0.265	0.236	0.262
	25%	0.542	0.148	0.141	0.137	0.166	0.164	0.099	0.097	0.104	0.106
	50%	0.301	0.122	0.115	0.113	0.118	0.034	0.086	0.084	0.084	0.085
	75%	0.276	0.141	0.134	0.129	0.139	0.073	0.090	0.089	0.084	0.078
	95%	0.507	0.187	0.182	0.168	0.260	0.161	0.132	0.132	0.111	0.107
(ii)	5%	1.214	0.488	0.453	0.379	0.432	0.312	0.273	0.273	0.243	0.235
	25%	0.493	0.172	0.135	0.144	0.157	0.160	0.085	0.086	0.090	0.080
	50%	0.384	0.177	0.124	0.129	0.131	0.061	0.087	0.088	0.099	0.103
	75%	0.329	0.211	0.162	0.186	0.197	0.054	0.158	0.152	0.206	0.217
	95%	0.309	0.202	0.181	0.166	0.235	0.162	0.106	0.104	0.088	0.103
(iii)	5%	1.727	0.266	0.266	0.444	0.836	0.196	0.225	0.218	0.319	0.364
	25%	0.916	0.250	0.280	0.248	0.392	0.232	0.202	0.205	0.266	0.256
	50%	0.317	0.134	0.156	0.226	0.245	0.098	0.153	0.152	0.312	0.293
	75%	0.132	0.143	0.162	0.142	0.103	0.026	0.116	0.113	0.110	0.128
	95%	0.173	0.170	0.185	0.345	0.308	0.099	0.116	0.118	0.418	0.321
(iv)	5%	1.140	0.557	0.572	0.373	0.393	0.325	0.357	0.356	0.203	0.236
	25%	0.532	0.193	0.216	0.222	0.175	0.212	0.182	0.184	0.263	0.229
	50%	0.212	0.117	0.140	0.126	0.117	0.038	0.075	0.077	0.097	0.095
	75%	0.474	0.136	0.161	0.254	0.349	0.191	0.084	0.084	0.312	0.286
	95%	2.067	0.517	0.439	0.963	1.691	0.667	0.504	0.438	0.843	0.948
(v)	5%	2.064	0.475	0.486	0.465	0.823	0.185	0.348	0.348	0.296	0.337
	25%	0.709	0.187	0.205	0.214	0.306	0.218	0.169	0.170	0.247	0.182
	50%	0.275	0.116	0.134	0.159	0.170	0.063	0.122	0.124	0.203	0.182
	75%	0.184	0.138	0.154	0.142	0.119	0.041	0.111	0.111	0.113	0.110
	95%	0.333	0.204	0.219	0.319	0.308	0.098	0.191	0.191	0.364	0.194

Table 6: AMSE and ABIAS of small area quantile estimators (Model B, $n = 1000$, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$).

		AMSE					ABIAS				
Error	α	Direct	EL(1)	EL(2)	NER	HNER	Direct	EL(1)	EL(2)	NER	HNER
	5%	0.264	0.066	0.064	0.064	0.077	0.145	0.107	0.107	0.107	0.106
	25%	0.079	0.031	0.029	0.029	0.033	0.026	0.058	0.056	0.057	0.054
(i)	50%	0.049	0.024	0.022	0.022	0.022	0.011	0.042	0.040	0.042	0.041
	75%	0.051	0.027	0.026	0.026	0.028	0.018	0.048	0.048	0.048	0.045
	95%	0.110	0.043	0.042	0.042	0.060	0.090	0.087	0.087	0.093	0.075
	5%	0.251	0.092	0.086	0.078	0.084	0.113	0.135	0.134	0.141	0.144
	25%	0.063	0.029	0.023	0.026	0.027	0.026	0.048	0.048	0.058	0.060
(ii)	50%	0.061	0.029	0.020	0.026	0.026	0.022	0.038	0.039	0.077	0.078
	75%	0.056	0.033	0.025	0.041	0.043	0.012	0.053	0.054	0.133	0.135
	95%	0.043	0.038	0.034	0.034	0.043	0.048	0.062	0.061	0.089	0.086
	5%	0.368	0.040	0.042	0.107	0.151	0.198	0.107	0.092	0.191	0.192
	25%	0.176	0.057	0.061	0.071	0.093	0.064	0.112	0.109	0.189	0.186
(iii)	50%	0.038	0.023	0.027	0.081	0.084	0.017	0.069	0.070	0.244	0.242
	75%	0.021	0.023	0.028	0.025	0.018	0.010	0.056	0.059	0.062	0.059
	95%	0.032	0.033	0.036	0.186	0.178	0.039	0.073	0.075	0.396	0.380
	5%	0.215	0.114	0.118	0.084	0.079	0.090	0.179	0.178	0.127	0.130
	25%	0.069	0.030	0.035	0.056	0.045	0.033	0.076	0.077	0.167	0.157
(iv)	50%	0.029	0.019	0.024	0.026	0.023	0.008	0.043	0.046	0.067	0.069
	75%	0.033	0.023	0.028	0.107	0.121	0.020	0.057	0.058	0.276	0.273
	95%	0.399	0.065	0.058	0.472	0.567	0.078	0.153	0.135	0.634	0.648
	5%	0.548	0.136	0.133	0.137	0.187	0.231	0.249	0.231	0.223	0.199
	25%	0.110	0.043	0.049	0.082	0.097	0.040	0.110	0.111	0.216	0.173
(v)	50%	0.038	0.020	0.025	0.053	0.054	0.012	0.044	0.048	0.168	0.161
	75%	0.029	0.030	0.035	0.037	0.023	0.010	0.077	0.079	0.106	0.065
	95%	0.079	0.073	0.076	0.206	0.165	0.075	0.176	0.175	0.388	0.303

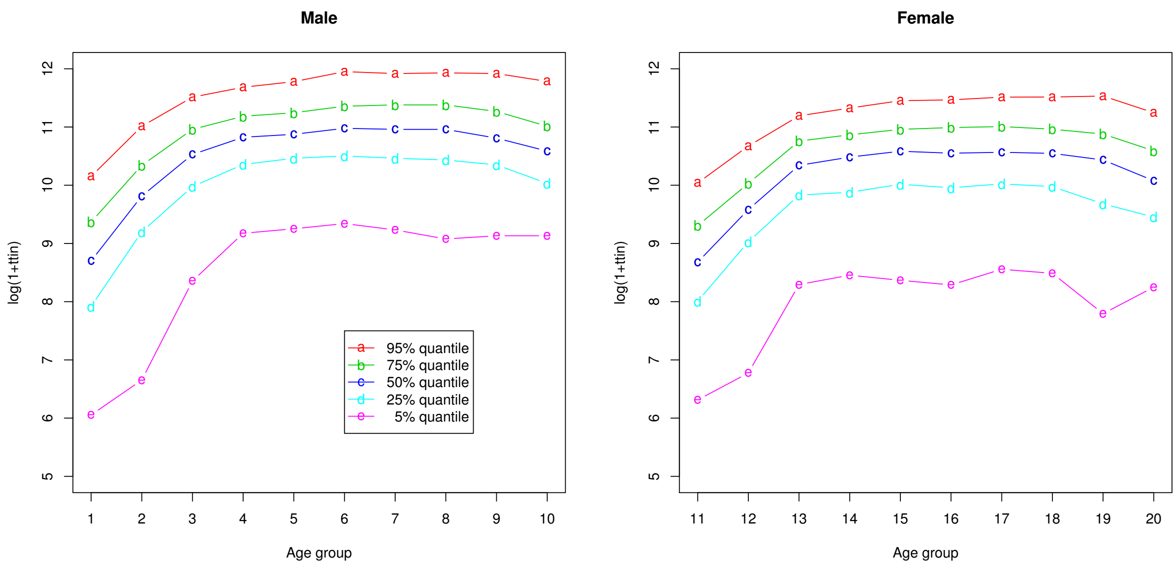

Figure 1: Small area population quantiles of the SLID data.

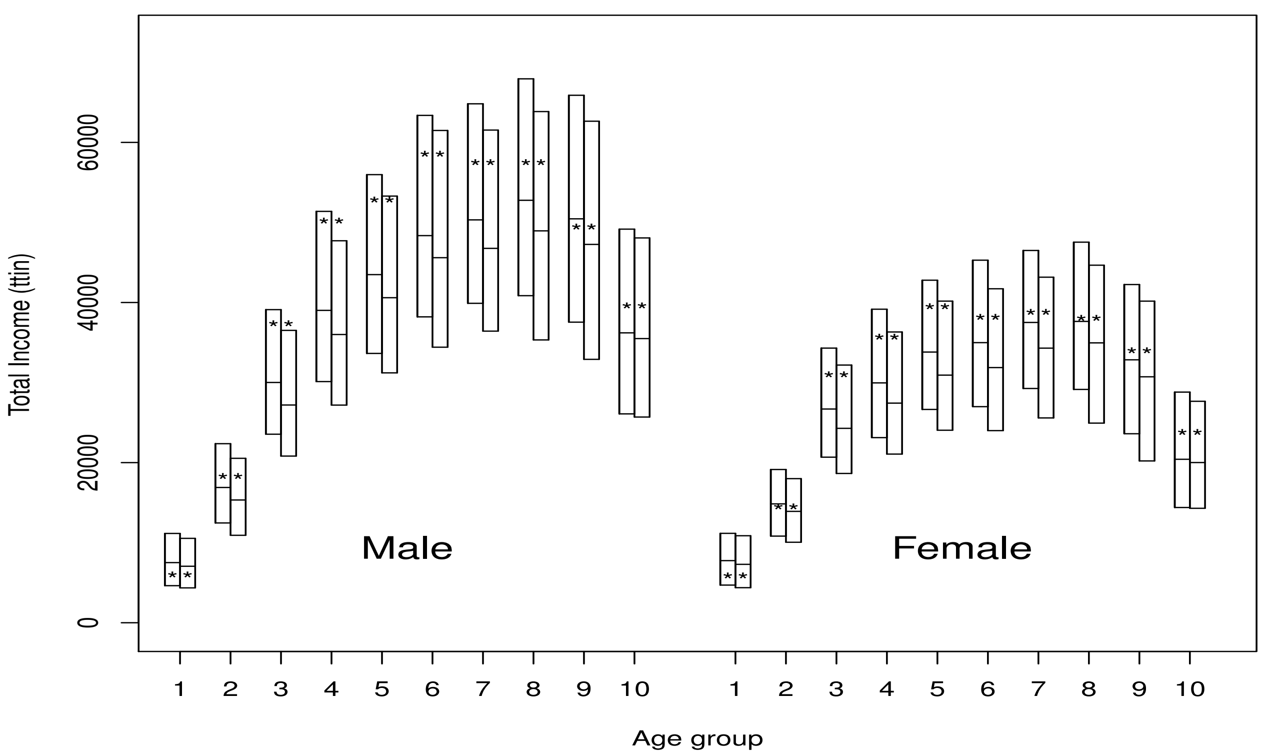

Figure 2: EL(1)/NER small area median estimation of the SLID population (n=200). $\star$: small area median; top, middle, and bottom lines: 90th, 50th, and 10th percentiles.

Table 7: AMSE and ABIAS of small area quantile estimators based on real data.

		AMSE					ABIAS				
n	α	Direct	EL(1)	EL(2)	NER	HNER	Direct	EL(1)	EL(2)	NER	HNER
	5%	1.765	0.453	0.442	0.421	0.649	0.195	0.417	0.408	0.397	0.365
	25%	0.340	0.099	0.103	0.173	0.196	0.179	0.208	0.209	0.340	0.251
200	50%	0.095	0.054	0.058	0.077	0.078	0.030	0.128	0.129	0.188	0.184
	75%	0.078	0.048	0.053	0.057	0.074	0.032	0.094	0.094	0.101	0.138
	95%	0.185	0.076	0.081	0.224	0.289	0.060	0.135	0.138	0.393	0.260
	5%	0.655	0.263	0.257	0.259	0.283	0.296	0.358	0.344	0.336	0.300
	25%	0.041	0.037	0.041	0.105	0.111	0.030	0.150	0.153	0.286	0.266
1000	50%	0.017	0.013	0.016	0.036	0.037	0.007	0.054	0.057	0.149	0.147
	75%	0.015	0.017	0.020	0.024	0.020	0.010	0.056	0.059	0.088	0.064
	95%	0.048	0.053	0.056	0.204	0.205	0.067	0.173	0.174	0.413	0.371