A RESTRICTED MODEL APPROACH TO IMPROVE THE PRECISION OF ESTIMATORS

CRISTINA RUEDA AND JOSÉ A. MENÉNDEZ

Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Universidad de Valladolid

C/ Prado de la Magdalena s/n, 47005 Valladolid, Spain

E-Mail: crueda@eio.uva.es

In this paper we propose a new approach to small area estimation that uses the methodology of constrained statistical inference (CSI) to improve the precision of direct estimates. The idea is to formulate a linking model for related domains using prior knowledge that is incorporated as restrictions on the model parameters.

The proposed estimators are indirect domain estimators and could be developed using explicit or implicit models, which will be called restricted models.

We focus on the estimation of a domain mean. Consider m domains and their means, $Y = (Y_1,..., Y_m)'$, as the parameters of interest and also consider the total mean $\overline{Y}$. The corresponding sample means based on a sample on each domain give the direct estimators, $\hat{Y} = (\hat{Y}_1,...,\hat{Y}_m)'$ and $\hat{\overline{Y}}$ respectively .

An example of a simple restricted model is given when the information $\overline{Y} \geq 0$ is included in the model. The corresponding restricted estimator is then $\hat{Y}^r = p(\hat{Y}/C_+)$, where $C_+ = \left\{ v \in R^m : \sum_{i=1}^{m} v_i \geq 0 \right\}$ and $p(y/C)$ is the projection of y onto a cone C . We introduce a more complex model when supplementary information in the form of an auxiliary variable $X = (X_1,...., X_m)'$ is available and a monotone relationship between X and Y exists, which can be formulated as follows: $X_i \leq X_j \Rightarrow Y_i \leq Y_j$. Statistically, this information is incorporated in the estimation process through the order relationship $\leq_x$ induced by X , the order cone $C_X = \left\{ v \in R^m : v_i \leq v_j \text{ if } i \leq_X j \right\}$ and the restriction $Y \in C_X$. The corresponding estimator is then $\hat{Y}^r = p(\hat{Y}/C_X)$. Intuitively, we would expect to do better by incorporating such additional information than by ignoring them.

In a similar way a restricted model could be defined using an explicit linear mixed constrained model given by $\hat{Y}_i = \mu_i + v_i + e_i$, $\mu \in C_X$. In this case a new restricted estimator would be obtained that have not been referenced before in the literature. In these and other similar models the properties of restricted and related estimators must be compared with classical alternatives to small area estimation.

The properties of the restricted estimator $\hat{Y}^r = p(\hat{Y}/C_X)$ have been extensively studied in the CSI literature, but as far as we know there have been no applications to small area problems. The more relevant result for the small area context is that using the criterion of the Mean Square Error (MSE), $\sum_{i=1}^{m} E(\hat{Y}_i^r - Y_i)^2$, the restricted estimator performs much better than the direct estimator when the hypothesis $Y \in C_X$ is true.

Taking into account constraints in the models produces a reduction of the parameter or sample space. We think that it would be possible to design methods that properly incorporate the constraints in the models, producing efficient estimators for small area applications.

In this first attempt we have considered the simplest model where only the information $\overline{Y} \geq 0$ is available.

We propose a family of estimators defined by $\hat{Y}^w = P(\hat{Y}/C_w)$ where C_w is a circular cone, $C_w = \left\{ v \in R^m : \langle c, v \rangle \leq \cos(w) \| v \| \right\}$, $\| c \| = 1$ and $w \in [0, \pi/2]$. Particular cases are the synthetic estimator, $\hat{Y}^{w=0} = \hat{\overline{Y}}$, and the restricted estimator, $\hat{Y}^{w=\pi/2} = p(\hat{Y}/C_+) = \hat{Y}^r$. In other cases $\hat{Y}^w$ could be considered as a kind of composite estimator. An *"empirical restricted"* estimator is selected from de above family in two steps. In the first step an optimum angle is defined by $\hat{w}_{opt} = \arg\min_w E(\hat{Y}^w - Y)^2$ and in the second step a plug-in estimator is obtained from the sample as $\hat{Y}^{\hat{w}_{opt}}$. In this paper we study some properties of this estimator and compare it with other classical counterparts as the positive part James-Stein estimator.