HANDBOOK OF STATISTICS
VOL 39

Sample Surveys: Theory, Methods and Inference

Editors: C.R. Rao and D. Pfeffermann (2009, forthcoming)

Chapter 31
Design-based methods of estimation for domains and small areas

Risto Lehtonen (University of Helsinki) and Ari Veijanen (Statistics Finland)

(ote kirjan luvusta 31)

2.2.5. Design-based properties of domain estimators

Known design-based properties related to bias, precision, and accuracy of model-assisted estimators are summarized in Table 1. For a comparison, properties of corresponding model-dependent estimators also are included in the table. Model-assisted estimators such as GREG are design consistent or nearly design unbiased by definition, but their variance can become large in domains where the sample size is small.

Model-dependent estimators such as synthetic and EBLUP estimators are design-biased: the bias can be large for domains where the model does not fit well. The variance of a model-dependent estimator can be small even for small domains, but the accuracy can be poor if the squared bias dominates the MSE, as shown, for example, by Lehtonen et al. (2003, 2005). For a model-dependent estimator, the dominance of the bias component together with a small variance can cause poor coverage rates and invalid design-based confidence intervals. For design-based model-assisted estimators, on the other hand, valid confidence intervals can be constructed.

Typically, model-assisted estimators are used for major or not-so-small domains, and model-dependent estimators are used for small domains where model-assisted estimators can fail. Table 1 indicates that small domains present problems in the design-based approach. Purcell and Kish (1980) call domain a mini domain when $N_d/N < 1\%$. In so small domains, especially, direct estimators can have large variance. Small domains are the main reason to prefer indirect model-based estimators to design-based estimators (Rao, 2005). By proper planning of the sampling strategy, it is possible to decrease the variance of a design-based estimator in the small domains. Singh et al. (1994) and Marker (2001) give examples of such strategies.

Table 1
Design-based properties of model-assisted and model-dependent estimators for domains and small areas

	Design-based model-assisted methods GREG and calibration estimators	**Model-dependent methods** Synthetic and EBLUP estimators
Bias	Design unbiased (approximately) by the construction principle	Design biased Bias does not necessarily approach zero with increasing domain sample size
Precision (Variance)	Variance may be large for small domains Variance tends to decrease with increasing domain sample size	Variance can be small even for small domains Variance tends to decrease with increasing domain sample size
Accuracy (Mean Squared Error, MSE)	MSE = Variance (or nearly so)	MSE = Variance + squared Bias Accuracy can be poor if the bias is substantial
Confidence Intervals	Valid design-based intervals can be constructed	Valid design-based intervals not necessarily obtained

Table 2. Application areas of estimation approaches by domain sample size

	DOMAIN SAMPLE SIZE		
ESTIMATION APPROACH	**Minor**	**Medium**	**Major**
Model-based			
Synthetic SYN	++	+	**0**
EBLUP	+++	++	++
Design-based			
Horvitz-Thompson HT	**0**	+	++
Model-assisted GREG	+	++	+++

Applicability

0 **Not at all**
+ **Low**
++ **Medium**
+++ **High**