

QR prediction for statistical data integration

Camelia Goga

Université de Bourgogne Franche Comté

In this paper, we investigate how a big non-probability database can be used to improve estimates of finite population totals from a small probability sample through data integration techniques. We study the class of QR predictors, introduced by Särndal and Wright (1984), to handle the less common case where the non-probability database contains no study variable but auxiliary variables. We also require that the non-probability database is large and can be linked to the probability sample. We provide conditions ensuring that the QR predictor is asymptotically design-unbiased. We derive its asymptotic design variance and provide a consistent design-based variance estimator. We compare the design properties of different predictors, in the class of QR predictors, through a simulation study. This class includes a model-based predictor, a model-assisted estimator, and a cosmetic estimator. In our simulation setups, the cosmetic estimator performed slightly better than the model-assisted estimator. These findings are confirmed by an application to La Poste data, which also illustrates that the properties of the cosmetic estimator are preserved irrespective of the observed non-probability sample.