Accounting for the Quality of Auxiliary Data Used in Small Area Estimation

John L. Eltinge
U.S. Bureau of Labor Statistics, Washington DC, USA,Eltinge.John@bls.gov

In work with small area estimation, preliminary exploratory work often identifies auxiliary data sources that could potentially provide valuable predictor variables. Examples include information from administrative records, transaction records and corporate data files. However, these data sources are often subject to important data-quality limitations, including incomplete-data patterns; cross-sectional or temporal aggregation effects; idiosyncrasies in operational definitions of variables; and measurement errors. Furthermore, at the initial evaluation stage, analysts often have limited empirical information on the magnitudes of the abovementioned problems, and on the practical impact that those problems would have on the properties of standard small area estimators. Following a review of applicable previous literature, this paper presents a suite of methods for evaluation of the prospective impact of data-quality problems on small area estimation. Special attention centers on (1) designs for preliminary assessment of the nature and magnitude of data-quality problems in candidate auxiliary data sources; (2) use of data from (1) to estimate relatively simple models for the identified quality problems; (3) use of the models developed in (2) to conduct sensitivity analyses for the resulting bias and variance of standard small area estimators that would use the candidate auxiliary variables; and (4) use of results from (3) to suggest alternative estimators that would account more fully for the predominant data quality problems identified in the previous steps. The proposed methods are applied to an example from the U.S. Bureau of Labor Statistics.

Keywords: non-designed data; organic data; sensitivity analysis; total survey error model.