Small area estimation by calibration methods

Risto Lehtonen*
University of Helsinki, Finland – risto.lehtonen@helsinki.fi

Ari Veijanen
Statistics Finland – ari.veijanen@stat.fi

We examine here the statistical properties (design bias and accuracy) of certain calibration methods for the estimation of finite population parameters (totals, means) for population subgroups or domains and small areas. Methods include model-free calibration and model calibration and a more recent method called "hybrid" calibration, a combination of model-free calibration and model calibration. Statistical properties of the methods are assessed by design-based simulation experiments.

In *model-free (or linear) calibration*, weights are calibrated to reproduce the known population or sub-population totals of the auxiliary variables (coherence criterion). Aggregate-level auxiliary data are used in calibration. No explicit model statement is required. The same set of calibrated weights can be supplied to the diverse survey variables. In official statistics production, these properties are often considered a benefit.

Model calibration represents a model-assisted technique, where an assisting model is explicitly stated. The weights are calibrated to reproduce the population total of the predictions derived via the specified model. In a model calibration procedure, the model is first fitted to the data and predictions are calculated to all population elements. In the calibration phase, calibration can be defined at the population level, at the domain level or at an intermediate level (for example at a neighbourhood that contains the domain of interest). Access to unit-level auxiliary data is assumed, and a considerable modelling effort is often needed. In the model calibration procedure, the method is applied separately for each survey variable. Coherence of the estimates of the auxiliary variables with published statistics is not guaranteed. Model calibration extends the scope of calibration beyond the linear model, to the members of generalized linear mixed models family.

"Hybrid" calibration combines some of the favourable properties of model-free calibration and model calibration. In this method, coherence property of estimates with published statistics is retained for a subset of auxiliary variables, and another subset is treated with model calibration involving flexible modelling. For example, some auxiliary variables (whose unit-level population data are available) are incorporated in the assisting model (to improve accuracy) and some variables (whose domain-level population totals are available) are incorporated in the model-free calibration procedure (for coherence with published statistics). The two sets of auxiliary variables can be distinct or they may overlap.

Keywords: Calibration techniques, Design-based methods, Generalized linear mixed models, Small area estimation.