Survey Inference Using Geocoded Data

Ray Chambers*
University of Wollongong, Australia - ray@uow.edu.au

Abstract

Geography has been part of sample surveys almost from the time that the survey method was accepted as the gold standard for scientific data collection and analysis, with geographic stratification and multilevel area sampling forming the backbone of sample design since the 1940s. However, the coarseness of the population location data provided by the maps used to implement these designs, and the cost of keeping them up to date, meant that there was little incentive to develop design-based survey estimation methods that made more intensive use of spatial dependence in a target population. Survey estimation methods effectively treated this dependence as just one of many contributing sources for overall sampling variability. Since 1990, however, we have seen a steady rise in the demand for survey estimates for small areas, with consequent general acceptance of model-based methods for survey inference. Coupled with the increasing availability of geocoded data down to quite fine spatial levels, this has driven development of survey estimation methods that make explicit use information about where a unit (e.g., person, household, farm, business) is located in the space of all units making up a target population as well as the more standard information about the values of the survey variables of interest and the values of auxiliary variables. In this presentation, I will review survey estimation methods that use location information to improve inference. In particular, I will contrast methods that exploit spatial correlation in model residuals with those that exploit spatial heterogeneity in model structure. Not surprisingly, small area estimation methods, and particularly those like poverty mapping where the spatial dimension is important, will feature prominently.

Keywords: Spatial data; Small area estimation; Poverty mapping; Spatial heterogeneity.