

Integration of Matrix Sampling and Multiple-Frame Methodology

John L. Eltinge

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

Some large-scale surveys use matrix sampling to reduce respondent burden. The main idea is that a given sample unit is only asked to provide responses for a subset of the items covered in the full survey instrument, with the subset determined through a randomization mechanism. In some cases, one may further reduce burden by obtaining some data through, e.g., administrative or transaction records. Such records have both strengths (e.g., coverage of some details that may be difficult to capture through standard interview processes) and limitations (e.g., failure to cover the full target population). This paper considers estimation and inference based on data from the resulting combined survey and administrative record sources. These methods require extension of ideas developed previously for matrix-sampling and multiple-frame-multiple-mode surveys. Three issues receive principal attention. First, the conceptual basis for estimation and inference depends on sources of random variability associated with the superpopulation considered to have generated the finite population of interest; the sampling mechanisms used for the formal survey design; and nonsampling error processes associated with incomplete data, measurement errors and aggregation effects encountered in the survey and administrative data. Second, efficient integration of the administrative data with the primary survey data generally will require information from supplementary surveys for, e.g., multiplicity measures, coverage assessment and regression coefficients for use in imputation, allocation and measurement error adjustments. Third, practical implementations of these ideas depend on available empirical information. This includes information on the quality of the data from each prospective source. This also includes empirical information on the fixed and marginal costs of specific survey sections under a matrix-sampling design, and costs associated with the capture and management of data from specific administrative record sources. Some ideas in this paper are motivated and illustrated by examples from the U.S. Consumer Expenditure Survey.

Key words: Administrative record data, data quality, respondent burden, survey cost structures