Dissertation Defense

Using LASSO to Calibrate Non-probability Samples using Probability Samples

Jack (Kuang-Tsung) Chen

Dissertation Committee: Mike Elliott (Chair), Rick Valliant, Sunghee Lee and Fred Feinberg

Wednesday, December 9th at 10:00am

368 ISR, University of Michigan
Via video to 1218 LeFrak, JPSM at University of Maryland

Amidst declining response rates and rapidly increasing cost of probability-based sampling, the resurgence of more cost-effective non-probability sampling has prompted survey researchers to explore different adjustment methods for non-probability samples. The current approach to post-survey adjustments attempts to create one single set of survey weights to correct all imbalances within a non-probability sample. A common scheme is to combine probability and non-probability samples to generate pseudo-selection-weights for non-probability sample respondents. The method requires a large probability-sampling-based data, and all variables related to propensity of a respondent being in the non-probability sample. In practice, obtaining an appropriate probability sample is too costly, and there is no systematic way to determine the correct variables that can generate weights to fix all errors of a non-probability sample. An alternative approach is to adjust the non-probability sample so that the weighted sample totals of a set of variables, known as calibration variables, equal to their Census benchmark totals. Although the method does not require specialized probability-sampling-based data, the set of calibration variables is small due to limited Census benchmark information. The resulting calibrated weights can only correct the imbalance with respect to the calibration variables, which is insufficient for adjusting all errors of a non-probability sample. To date, no method has shown to be effective in helping researchers make unbiased inference from non-probability samples. This dissertation addresses the growing demand for making proper inference from non-probability samples. Instead of generating a single set of weights to fix all errors in a non-probability sample, we focus on constructing weights to enable unbiased inference for a specific outcome of interest. We introduce the Least Angle Shrinkage and Selection Operator, LASSO, to the framework of model-assisted calibration. The proposed method, LASSO calibration, determines the set of variables with the strongest relation to the outcome variable, then estimates the expected population total of the outcome based on a probability benchmark sample. The calibrated weights adjust the outcome variable in the non-probability sample to match the expected outcome in the population. The estimator of population total based on LASSO calibrated weights maintains good frequentist properties in terms of bias and variance. The theoretical framework is developed and evaluated through simulations. An application of LASSO calibration to a large-scale internet-based non-probability sample shows the proposed method can make more accurate and precise inference than existing methods.