Proceedings of Statistics Canada Symposium 2018
Combine to Conquer: Innovations in the Use of Multiple Sources of Data

Calibrating Non-Probability Samples With Probability Samples Using LASSO

Michael R. Elliott, Jack K-T Chen, and Richard L. Valliant[1]

Abstract

Due to declining telephone survey response rates, it has become challenging for election pollsters to capture voting intentions in a timely way. This has lead to the expanded use of samples obtained from non-probability web surveys. Because non-probability samples can suffer from selection bias, we develop a model-assisted calibration method using adaptive LASSO regression – estimated-controlled LASSO (ECLASSO). This method yields consistent estimates of population totals as long as a subset of the true predictors is included in the prediction model, thus allowing large numbers of possible covariates to be included without risk of overfitting. We apply ECLASSO to predict the voting results for the U.S. 2014 midterm election.

Key Words: Probability survey; Propensity weighting; General regression estimator; Model-assisted calibration; Election polls.

1. Introduction

Non-probability samples are an increasing part of life for the survey analyst. This is due to several reasons. Declining land-line and improved telephone screening technology has lead to major problems with the use of telephone survey to capture voting intentions in a timely manner (Kohut et al. (2012), Sturgis et al. (2016)). Increasing levels of non-response (Dutwin and Lavrakas 2016) and increasing costs pose challenges as well. On the positive side, non-probability samples can provide detailed measures of interest not present in probability samples, as well as larger sample sizes for less cost, especially in small domains. This offers the possibility of improved inference if increases in precision are not overwhelmed by selection bias from the non-probability sample.

2. A Framework for Nonprobability Sample Inference

The resurgence of non-probability sampling has prompted survey researchers to explore different adjustment methods for non-probability samples using probability samples. Elliott and Valliant (2017) review work in this area, dividing methods into "quasi-likelihood" approaches (Schonlau et al., 2004) versus "superpopulation" modeling approaches (Valliant et al., 2000). Consider the joint density of a population vector of analysis variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ and of 0-1 indicator variables for a sample s. In the most general setting this density can be factored as

$$f(\mathbf{Y}, \boldsymbol{\delta}_s | \mathbf{X}; \Theta, \Phi) = f(\mathbf{Y} | \mathbf{X}; \Theta) f(\boldsymbol{\delta}_s | \mathbf{Y}, \mathbf{X}; \Phi)$$

where $\mathbf{X}$ is an $N \times p$ matrix of covariates that govern $\mathbf{Y}$ through an unknown parameter Θ, and an unknown parameter Φ governs $f(\boldsymbol{\delta}_s)$ through both $\mathbf{Y}$ and $\mathbf{X}$ (Smith 1983; Rubin 1976; Little 1982).

In probability sampling, the sampling indicator depends only on $\mathbf{X}$ though known parameters: $f(\boldsymbol{\delta}_s | \mathbf{Y}, \mathbf{X}; \Phi) = f(\boldsymbol{\delta}_s | \mathbf{X})$; the resulting known probabilities can be used to produce sampling weights. In non-probability sampling, $\boldsymbol{\delta}_s$ can depend on $\mathbf{Y}$ and/or Φ in addition to $\mathbf{X}$; assuming that sufficient $\mathbf{X}$ are available so that $\boldsymbol{\delta}_s \perp \mathbf{Y} \mid \mathbf{X}$, $f(\boldsymbol{\delta}_s | \mathbf{X}; \Phi)$ can be modeled and

[1]Michael R. Elliott, University of Michigan Institute for Social Research, 426 Thompson St., Ann Arbor, Michigan, USA 48106 (mrelliot@umich.edu); Jack K-T Chen, Survey Monkey, Palo Alto, California, USA, 94301 (jjkkttcc@gmail.com); Richard L. Valliant, University of Michigan Institute for Social Research, 426 Thompson St., Ann Arbor, Michigan, USA 48106 (valliant@umich.edu)

the resulting estimated probabilities inverted to obtain 'pseudo-weights'; this is sometimes termed propensity score weighting or *quasi-randomization*. An alternative approach, *superpopulation modeling*, focuses on the underlying relationship between **Y** and **X** by modeling $f(\mathbf{Y}|\mathbf{X};\Theta)$. By partitioning **Y** into sampled and non-sampled units, $f(\mathbf{Y} \mid \mathbf{X};\boldsymbol{\Theta}) = f(\mathbf{Y}_s \mid \mathbf{Y}_{\bar{s}}, \mathbf{X};\boldsymbol{\Theta})\, f(\mathbf{Y}_{\bar{s}} \mid \mathbf{X};\boldsymbol{\Theta})$,we can use model estimates from the sample to predict non-sampled elements under the assumption that $f(\mathbf{Y}_s \mid \mathbf{Y}_{\bar{s}}, \mathbf{X};\boldsymbol{\Theta}) = f(\mathbf{Y}_s \mid \mathbf{X};\boldsymbol{\Theta})$.

3. Calibration

3.1 Generalized Regression Estimation

Starting from the assumptions that a probability sample is used and that the (model) expectation of **Y** is linear in **X**

$$E_M(Y_i) = \mathbf{X}_i^T \boldsymbol{\beta} \tag{1}$$

for an unknown parameter $\boldsymbol{\beta}$ of length *p*, we can solve an estimating equation for $\boldsymbol{\beta}$ using sample data to obtain the least squares estimator

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_s^T \mathbf{D} \mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \mathbf{D} \mathbf{Y}_s$$

where $\mathbf{D} = diag(d_i)$ for design weights d_i. We can then predict the non-sampled elements $\hat{\mathbf{Y}}_{\bar{s}}$ using $\hat{\mathbf{Y}}_{\bar{s}} = \mathbf{X}_{\bar{s}} \hat{\boldsymbol{\beta}}$ where $\mathbf{X}_{\bar{s}}$ is the $(N-n)\times p$ matrix of auxiliaries for the nonsample units. A predictor of the population total is then given by

$$\hat{T} = \sum_{i \in s} w_i Y_i = \sum_{i \in s} d_i Y_i + \left(\mathbf{T}_{Ux} - \hat{\mathbf{T}}_x\right) \hat{\boldsymbol{\beta}} \tag{2}$$

where $\mathbf{w} = \mathbf{d} + \left(\mathbf{T}_{Ux} - \mathbf{T}_{sx}\right)^T \left(\mathbf{X}_s^T \mathbf{D} \mathbf{X}_s\right)^{-1} \mathbf{X}^T \mathbf{D}$, $\mathbf{T}_{sx} = \sum_{i \in s} d_i \mathbf{x}_i$. This corresponds to the *generalized regression estimator* or GREG of Deville and Särndal (1992). (If **X** is an *H*-level categorical variable, then $\hat{T}$ corresponds to the stratified estimator: $\hat{T}^{ps} = \sum_{h=1}^{H} N_h \bar{y}_{sh}$, where N_h corresponds to the sample size in the *h*th stratum and $\bar{y}_{sh}$ to the sample mean in the *h*th stratum.) This estimator is model-unbiased under (1) and approximately design-unbiased if a probability sample is selected with probabilities given by d_i^{-1}.

The development above assumes that design weights are available as they would be in a probability sample. If the sample is non-probability, then **D** is omitted, and the population total can be estimated by the prediction estimator

$$\hat{T} = \sum_{i \in s} Y_i + \left(\mathbf{T}_{Ux} - \mathbf{T}_{sx}\right) \hat{\boldsymbol{\beta}}$$

with $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_s^T \mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \mathbf{Y}_s$ and $\mathbf{T}_{xs} = \sum_{i \in s} \mathbf{x}_i$. This estimator is model-unbiased under (1). An alternative would be to set $\mathbf{D} = diag\left(N/n\right)$ and use (2), as is done here in section 4.

In many cases, the availability of known control totals for the entire population $\mathbf{T}_{Ux}$ may be limited, making the assumption that $f(\mathbf{Y}_s \mid \mathbf{Y}_{\bar{s}}, \mathbf{X};\boldsymbol{\Theta}) = f(\mathbf{Y}_s \mid \mathbf{X};\boldsymbol{\Theta})$ rather strong. In this case, there may be a "benchmark" probability survey available with a richer set of **X** covariates available. We replace $\mathbf{T}_{Ux}$ in (2) with $\mathbf{T}_{Bx}$, the population total

estimated from the probability survey (Dever and Valliant 2010); we refer to the resulting estimators as estimated control generalized regression estimators, or ECGREG estimators.

3.2 Model-Assisted Calibration

The weights w_i in GREG can be viewed (Deville and Särndal 1992) as the weights that minimize $\sum_{i\in s}\left(w_i - d_i\right)^2 / d_i$ subject to the constraint that $\sum_{i\in s} w_i \mathbf{x}_i = \mathbf{T}_{Ux}$ (for standard calibration to known control totals) or $\sum_{i\in s} w_i \mathbf{x}_i = \mathbf{T}_{Bx}$ (for calibration to estimated benchmark control totals). *Model-assisted calibration* (Wu and Sitter 2001) minimize $\sum_{i\in s}\left(w_i - d_i\right)^2 / d_i$ subject instead to the constraints that $\sum_{i\in s} w_i = N$ and $\sum_{i\in s} w_i \hat{y}_i = \sum_{i\in U} \hat{y}_i$. Model-assisted calibration weights are given by $\mathbf{w}^{MC} = \mathbf{d} + \mathbf{DM(M}^T\mathbf{DM)}^{-1}\mathbf{(T - d}^T\mathbf{M)}^T$, where $\mathbf{T} = \left(N, \sum_{i\in U}\hat{Y}_i\right)$ and $\mathbf{M} = \left(\mathbf{1}_{n_s}, \left(\hat{Y}_i\right)_{i\in s}\right)$, and $\mathbf{1}_{n_s}$ is a vector of 1s of length equal to the sample size. It can be shown that the model-assisted estimator of the total given by the weighted sum of the model-assisted weights is given by

$$\hat{T}^{MC} = \sum_{i\in s} w_i^{MC} Y_i = \sum_{i\in s} d_i Y_i + \left(\sum_{i\in U}\hat{Y}_i - \sum_{i\in s} d_i \hat{Y}_i\right)\hat{B} \tag{3}$$

where the $\hat{B}$ that satisfies the calibration constraints is given by the design-weighed correlation between the observed and predicted values:

$$\hat{B} = \frac{\sum_{i\in s} d_i \left(\hat{Y}_i - \bar{\hat{Y}}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i\in s} d_i \left(\hat{Y}_i - \bar{\hat{Y}}\right)^2}$$

where $\bar{\hat{Y}} = \sum_{i\in s} d_i \hat{Y}_i \Big/ \sum_{i\in s} d_i$ and $\bar{Y} = \sum_{i\in s} d_i Y_i \Big/ \sum_{i\in s} d_i$ are the design-weighted means of the predicted and observed values, respectively. As long as the original design weights produce unbiased estimates, $\hat{T}^{MC}$ is approximately design unbiased when the sample size is large; the "assisting model" given by (1) improves efficiency to the extent that it is accurate. Just as with ECGREG estimators, we can replace covariates obtained from a population with covariates estimated from a benchmark sample s_B (we now denote our primary analytic sample by s_A). This replaces population totals $\mathbf{T}$ with $\hat{\mathbf{T}} = \left(\sum_{i\in s_B} d_i^B, \sum_{i\in s_B} d_i^B \hat{Y}_i\right)$, and our estimator of the population total (3) is updated as

$$\hat{T}^{ECMC} = \sum_{i\in s} w_i^{ECMC} Y_i = \sum_{i\in s_A} d_i^A Y_i + \left(\sum_{i\in s_B} d_i^B \hat{Y}_i - \sum_{i\in s_A} d_i^A \hat{Y}_i\right)\hat{B} \tag{4}$$

which we refer to as estimated control model-assisted calibrated, or ECEM, estimators.

3.3 ECLASSO

Here we develop estimated control lasso calibration, or ECLASSO. In many cases we may want to use a large vector of potential control totals for model (1), particularly if we are obtaining them from a benchmark probability survey. However, use of a large number of covariates might yield unstable prediction; hence, instead of obtaining $\hat{\boldsymbol{\beta}}$ by least squares as in ECEM, we use adaptive LASSO (Zhou 2006), a more robust estimation procedure. McConville et al (2017) and Chen et al. (2018) considered the use of adaptive LASSO for model-assisted calibration in setting with probability sampling or known population values; here we consider its use when the analytic sample is a non-probability sample and the benchmark sample is a probability sample.

The adaptive LASSO regression coefficients are obtained by solving a penalized regression equation. For linear adaptive LASSO regression, this is

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta}\left(\sum_{i\in s_A}\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\left|\hat{\beta}_j^{MLE}\right|^{-\gamma} \right)$$

For logistic adaptive LASSO, this is

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta}\left(\sum_{i\in s_A}\left[-y_i\left(\mathbf{x}_i^T\boldsymbol{\beta}\right) + \log\left(1+\exp(\mathbf{x}_i^T\boldsymbol{\beta})\right)\right] + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\left|\hat{\beta}_j^{MLE}\right|^{-\gamma} \right)$$

The factor $1/\left|\hat{\beta}_j^{MLE}\right|^{\gamma}$ balances the selection of covariates with large effect sizes in favor of lowering prediction error when the sample size is small. Once values of λ and γ are fixed, $\hat{\boldsymbol{\beta}}$ can be calculated through iterative procedures (Friedman et al. 2010); these algorithms are implemented in the R package *glmnet*. Values of λ and γ can be explored over a grid of values and selected via cross-validation. Adaptive LASSO has a model-consistency property known as the oracle property, which states that, under the condition that there is a regression model in which the parameters have both non-zero $\boldsymbol{\beta}^{(1)}$ and zero components $\boldsymbol{\beta}^{(2)}$ and λ grows at least at the rate of $\sqrt{n}/\left(\sqrt{n}\right)^{\gamma}$ but not faster than $\sqrt{n}$, $P(\hat{\boldsymbol{\beta}}^{(2)}=0)\to 1$ and $\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{(1)}-\boldsymbol{\beta}^{(1)}\right) \to N(\mathbf{0},\mathbf{C})$ where $\mathbf{C} = I^{-1}(\boldsymbol{\beta}^{(1)})$ is the inverse of the Fisher information matrix of $\boldsymbol{\beta}$. Said less technically, subject to regularity conditions, adaptive LASSO will converge to the true model as long as all true linear predictors are in the model.

Having obtained $\hat{\boldsymbol{\beta}}$ via adaptive LASSO, we obtain $\hat{Y}_i = \mathbf{x}_i^T\boldsymbol{\beta}$ and compute $\hat{T}^{ECLASSO}$ as in (4). It can be shown that, as long as the benchmark sample has the correct design weights $\hat{T}^{ECLASSO}$ will be asymptotically design and model unbiased, with an asymptotic design variance given by

$$v_{\mathrm{A}}(\hat{T}^{ECLASSO}) = \sum_{i\in s_A}\left(\frac{y_i-\hat{Y}_i\hat{B}}{\pi_i^A}\right)^2(1-\pi_i^A) + \sum_{i\in s_A}\sum_{j\neq i}\frac{\pi_{ij}^A-\pi_i^A\pi_j^A}{\pi_{ij}^A}\frac{(Y_i-\hat{Y}_i\hat{B})}{\pi_i^A}\frac{(Y_j-\hat{Y}_j\hat{B})}{\pi_j^A} +$$
$$\sum_{i\in s_B}\left(\frac{\hat{Y}_i\hat{B}}{\pi_i^B}\right)^2(1-\pi_i^B) + \sum_{i\in s_B}\sum_{j\neq i}\frac{\pi_{ij}^B-\pi_i^B\pi_j^B}{\pi_{ij}^B}\frac{\hat{Y}_i\hat{B}}{\pi_i^B}\frac{\hat{Y}_j\hat{B}}{\pi_j^B} \qquad \backslash$$

When the analytic sample A is a non-probability sample, we set $\pi_i^A \equiv \frac{n_A}{N}$ and $\pi_{ij}^A = \frac{n_A(n_A-1)}{N(N-1)}$. An alternative approach is to use a bootstrap estimator, which can be obtained by drawing one finite-population bootstrap of the benchmark sample and one simple-random-sample with replacement from the analytical sample, and then compute $\hat{T}^{ECLASSO}$ for each bootstrap sample. Simulation studies (not shown) find the analytic variance tends to underestimate the true variance, whereas a bootstrap estimator tends to be conservative.

4. Predicting 2014 US Senate and Governors Races

A random 10% of respondents who completed a SurveyMonkey poll in October 2014 were asked to provide voting preferences in Senate and governor races; approximately 2-3% did so. Although the sample was randomly selected among the survey takers, the response rate was low and, more importantly, the pool of respondents who completed an

initial SurveyMonkey survey is non-probability-based and may not be representative of the voting population. Because conditioning on likely voters improves election prediction (Bolstein, 1991; Gutsche et al., 2014), we restricted our analysis to those who indicated they: (1) already voted, (2) were absolutely certain to vote, or (3) were very likely to vote, and further who indicated they would vote for either a Democratic or Republican candidate, the two major US political parties. When further restricted to 8 states with Senate races (GA, IL, MI, MN, NJ, NC, TX, VA) and 11 states with gubernatorial races (AZ, CA, FL, GA, IL, MI, NY, OH, PA, TX, WI) that also had sufficient sample sizes of benchmark survey samples, the final analytical sample sizes are 33,199 for the collection of governor races and 28,686 for the collection of Senate races.

The benchmark probability sample of potential voters was obtained via telephone and cellphone during September and October of 2014 by the Pew Research Center (http://www.pewresearch.org). The benchmark sample consisted of 1,094 gubernatorial voters and 656 Senatorial voters. The surveys contained a large set of common covariates, including: age, gender, race, education, religion, religious attendance, approval of Obama, party preference.

The target of inference is the voting spread between Republicans and Democrats in state *r*, defined as

$$\hat{S}_{R(r)-D(r)} = \sum_{i \in s_{A(r)}} w_i y_i \Big/ \sum_{i \in s_{A(r)}} w_i - \sum_{i \in s_{A(r)}} w_i (1-y_i) \Big/ \sum_{i \in s_{A(r)}} w_i = 2 \sum_{i \in s_{A(r)}} w_i y_i \Big/ \sum_{i \in s_{A(r)}} w_i - 1$$

where $s_{A(r)}$ is the set of respondents in state *A*, y_i is an indicator for favoring the Republican candidate, and w_i depends on the estimator considered. In particular, we consider 4 estimators:

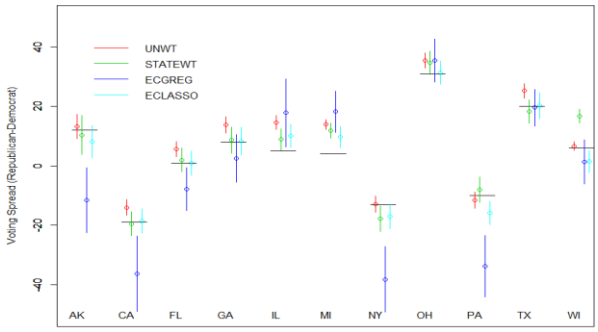
- UNWT: Unadjusted ($w_i = 1$).
- STATEWT: Calibrated to state-level measures from probability survey ($w_i = \hat{p}_i(\mathbf{x}_i)^{-1}$ where $\hat{p}_i(\mathbf{x}_i) = \exp\left(\mathbf{x}_i^T \hat{\boldsymbol{\alpha}} \big/ (1+\mathbf{x}_i^T \hat{\boldsymbol{\alpha}})\right)$ and $\hat{\boldsymbol{\alpha}}$ is estimated from a logistic regression of an indicator for being in the benchmark survey of covariates).
- ECGREG: Model assisted-calibration using GREG (w_i as defined in Section 3.2).
- ECLASSO: Model assisted-calibration using LASSO (w_i as defined in Section 3.3).

All confidence intervals were computed using bootstrap. Because the true results of the election are known, we can compute estimates of bias, root mean square error (RMSE), and coverage. Given the limited number of replications (11 gubernatorial races and 8 Senate races), we considered an α-level of 0.20 rather than 0.05 for coverage.

4.1 Results: gubernatorial races

Figure 1 shows results for 11 governor election forecasts. UNWT, STATEWT, and ECLASSO predicted the correct winning political party for all states in the analysis. ECGREG predicted Arizona and Florida incorrectly. However, without weighting adjustments, the sample has Republican overrepresentation, with 10 out of 11 states biasing toward Republican candidates. STATEWT reduced the bias for most states, while ECGREG appears to have over-adjusted toward Democratic direction. ECLASSO reduced unadjusted absolute sample bias to a maximum of 6% of true values across the 11 states, versus 10%-25% for the other estimators. On average, ECLASSO also has the smallest relative error across the states (0.5% D versus 1.9% R to 7.0% D for the other estimators), as well as the smallest RMSE (4.7% vs. 5.2% to 15.0% for the other estimators). Finally, STATEWT and ECLASSO had the best interval coverage (7 out of 11 or 64% of the 80% CIs).

Figure 4.1-1
Results for gubernatorial races: point estimates and 80% confidence intervals

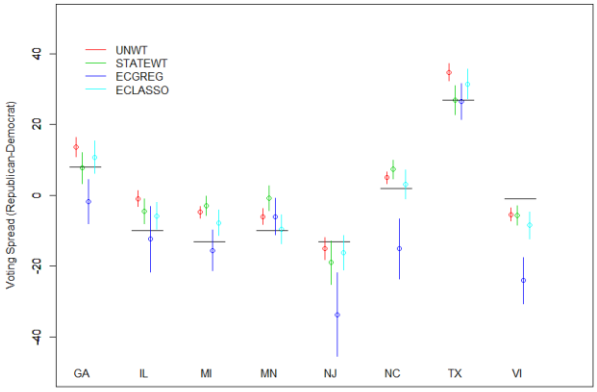


Note: Black line shows actual voting spread.

4.2 Results: Senate races

Figure 2 lists results for 8 Senate election forecasts. UNWT, STATEWT, and ECLASSO predicted the correct winning political party for all states in the analysis. ECGREG predicted Georgia and North Carolina incorrectly. Similar to the governor sample, the Senate sample has more Republican votes than the true voting spread, with 6 out of 8 states biasing toward Republican candidates. STATEWT reduced the bias for the majority of states, while ECGREG overadjusted in the Democratic direction. ECLASSO reduced unadjusted absolute sample bias to a maximum of 8% of true values across the 8 states, versus 9%-27% for the other estimators. On average, ECLASSO also has the smallest relative error across the states (1.0% R versus 2.4% R to 9.0% D for the other estimators), as well as the smallest RMSE (5.1% vs. 6.0% to 12.2% for the other estimators). Here ECLASSO was again tied for the best interval coverage (4 out of 8 or 50% of the 80% CIs), but this time with ECGREG (whose intervals were much wider).

Figure 4.2-1
Results for Senate races: point estimates and 80% confidence intervals



Note: Black line shows actual voting spread.

5. Discussion and Next Steps

Calibration is an important method to consider to deal with selection bias in non-probability samples. Here we developed estimated control lasso calibration, or ECLASSO, which uses adaptive LASSO to leverage large numbers of potential covariates in benchmark probability surveys to adjust estimates obtained from non-probability surveys using model-assisted calibration. We show that a relatively small benchmark sample can be used to substantially improve estimation of 2014 US governor and Senate races that relied on a non-probability poll of SurveyMonkey users, both in comparison to unadjusted estimates, as well as estimates obtained using alternative methods such as propensity score adjustment or generalized regression estimation.

Combining data from probability and non-probability samples remains an open area for research: use of propensity score, mean or quantile matching, mode effects, measurement error, and data harmonization and alignment are all important topics. We hope the application discussed here will encourage such work.

References

Bolstein, R. (1991), "Predicting the likelihood to vote in pre-election polls", *The Statistician*, 27, pp. 277–283.

Chen, J.K.T., R. L. Valliant, and M. R. Elliott (2018), "Model-assisted calibration of non-probability sample survey data using adaptive LASSO", *Survey Methodology*, 44, pp. 117-144.

Dever, J., and R. L. Valliant (2010), "A comparison of variance estimators for poststratification to estimated control totals", *Survey Methodology*, 36, pp. 45–56.

Deville, J. C., and C. E. Särndal (1992), "Calibration estimators in survey sampling", *Journal of the American Statistical Association*, 87, pp. 376-382.

Dutwin, D., and P. Lavrakas (2016), "Trends in telephone outcomes", *Survey Practice*, 9(2), pp. 1–9.

Elliott, M.R., and R. L. Valliant (2017), "Inference for non-probability samples", *Statistical Science*, 32, pp. 249–264.

Friedman, J., T. Hastie, and R. Tibshirani (2010), "Regularization paths for generalized linear models via coordinate descent", *Journal of Statistical Software*, 33, pp. 1–22.

Gutsche, T., A. Kapteyn, E. Meijer, and B. Weerman (2014), "The RAND continuous 2012 presidential election poll", *Public Opinion Quarterly*, 78, pp. 233–254.

Kohut, A., S. Keeter, C. Doherty, M. Dimock, and L. Christian (2012), *Assessing the representativeness of public opinion surveys.*

Little, R. J. A. (1982), "Models for nonresponse in sample surveys", *Journal of the American Statistical Association*, 77, pp. 237–250.

McConville, K., F. Bredit, T. Lee, and G. Moisen (2017), "Model-assisted survey regression estimation with the lasso", *Journal of Survey Statistics and Methodology*, 5, pp. 131–158.

Rubin, D. (1976), "Inference and missing data", *Biometrika*, 63, pp. 581–592.

Schonlau, M., K. Zapert, L. Simon, K. Sanstad, S. Marcus, J. Adams, M. Spranca, H. Kan, R. Turner, and S. Berry (2004), "A comparison between responses from a propensity weighted web survey and an identical RDD survey", *Social Science Computer Review*, 22, pp. 128–138.

Smith, T.M. F. (1983), "On the validity of inferences from non-random samples", *Journal of the Royal Statistical Society*, A146, pp. 394–403.

Sturgis, P., N. Baker, M. Callegaro, S. Fisher, J. Green, W. Jennings, J. Kuha, B. Lauderdale, and P. Smith (2016), *Report of the Inquiry into the 2015 British general election opinion polls*.

Valliant, R.L., A. H. Dorfman, and R. M. Royall (2000), *Finite Population Sampling and Inference: A Prediction Approach*, New York: Wiley.

Wu, C., and R. Sitter (2001), "A model-calibration approach to using complete auxiliary information from survey data", *Journal of the American Statistical Association*, 96, pp. 185–193.

Zou, H. (2006), "The adaptive lasso and its oracle properties", *Journal of the American Statistical Association*, 101, pp. 1418–1429.