Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Model-assisted calibration of non-probability sample survey data using adaptive LASSO

SURVEY
METHODOLOGY

A JOURNAL
PUBLISHED BY
STATISTICS CANADA

Canada

by Jack Kuang Tsung Chen, Richard L. Valliant and Michael R. Elliott

Release date: June 21, 2018

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-877-287-4369

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2018
Vol. 44, No. 1, pp. 117-144
Statistics Canada, Catalogue No. 12-001-X

Model-assisted calibration of non-probability sample survey data using adaptive LASSO

Jack Kuang Tsung Chen, Richard L. Valliant and Michael R. Elliott[1]

Abstract

The probability-sampling-based framework has dominated survey research because it provides precise mathematical tools to assess sampling variability. However increasing costs and declining response rates are expanding the use of non-probability samples, particularly in general population settings, where samples of individuals pulled from web surveys are becoming increasingly cheap and easy to access. But non-probability samples are at risk for selection bias due to differential access, degrees of interest, and other factors. Calibration to known statistical totals in the population provide a means of potentially diminishing the effect of selection bias in non-probability samples. Here we show that model calibration using adaptive LASSO can yield a consistent estimator of a population total 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 show that the model calibration using adaptive LASSO provides improved estimation with respect to mean square error relative to standard competitors such as generalized regression (GREG) estimators when a large number of covariates are required to determine the true model, with effectively no loss in efficiency over GREG when smaller models will suffice. We also derive closed form variance estimators of population totals, and compare their behavior with bootstrap estimators. We conclude with a real world example using data from the National Health Interview Survey.

Key Words: Adaptive LASSO estimators; Generalized regression estimator; Non-representative sample; Over-fitting; Variable selection; Oracle property.

1 Introduction

Probability-based sampling has dominated survey research for the greater part of the past century (Stephan, 1948; Frankel and Frankel, 1987). Given complete measures on sampled units with known selection probabilities, randomization theory removes selection bias by generating representative samples of the target population. On the other hand, non-probability samples generated without known selection probabilities are automatically at risk for selection bias, as samples can differ from the target population on key statistics (Groves, 2006). Well-documented failures in 1936 and 1948 presidential election polls highlight the potential downfalls in making population inference from non-probability samples (Mosteller, 1949).

Although the probability-sampling-based framework provides survey practitioners precise mathematical tools to assess and correct sampling errors, declining response rates among traditional data collection methods raise concerns over the potentially high nonresponse bias of probability samples. Pew Research reported that their response rates (RRs) in typical telephone surveys dropped from 36% in 1997 to 9% in 2012 (Kohut, Keeter, Doherty, Dimock and Christian, 2012), suggesting that telephone-based probability sampling may no longer be a viable methodology for general population surveys. In addition, obtaining data without exercising much control over the set of units for which it is collected is often cheaper and quicker than probability sampling. For these reasons non-probability sampling is currently staging a kind of

[1] Jack Kuang Tsung Chen is Statistician, Survey Monkey Inc., Palo Alto, CA. E-mail: jjkkttcc@gmail.com; Richard L. Valliant is Research Professor, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI. E-mail: valliant@umich.edu; Michael R. Elliott is Research Professor, Survey Research Center, Institute for Social Research, and Professor, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI. E-mail: mrelliot@umich.edu.

renascence (Baker, Brick, Bates, Battaglia, Couper, Dever, Gile and Tourangeau, 2013; Elliott and Valliant, 2017). Online data collection, a platform without a universal sampling frame to conduct probability-based sampling, was estimated to comprise nearly half of all U.S. survey research spending in 2012 (Terhanian and Bremer, 2012), and has almost certainly grown since then.

For many survey agencies, adjusting survey weights to known auxiliary information is the final and most crucial step in the weight construction process. Standard approaches include poststratification, in which weights are adjusted so that the weighted sample distribution of categorical auxiliary variables matches that of the population, and its extention to generalized regression estimation (GREG), which ensures that the weighted sum of each auxiliary variable (continuous or categorical) equals to its corresponding total in the population (Deville and Särndal, 1992). Calibration plays an important role in official statistics because it can generate weights such that the weighted demographic estimates across different surveys are consistent.

In probability samples, when design weights are equal to the inverse of selection probabilities, weighted estimates of sample totals are design-unbiased for the population total. Calibration adjusts design weights by a minimal degree so that the weighted sample totals for auxiliary variables match their known population totals (Särndal, Swensson and Wretman, 1992). In the probability sampling setting, calibration is introduced to reduce variance and/or correct for bias by adjusting for undercoverage or overcoverage of sub-groups of the sample. For large samples, the final calibrated weights can be applied to all variables in the survey, because they approximately maintain the unbiased property of original design weights. In non-probability samples, however, there are no selection probabilities to construct initial design weights that can produce unbiased estimates. Thus, there is no guarantee that the traditional calibrated weights can work for all variables in the non-probability sample. To make inference from non-probability samples, one practical approach is to construct a set of weights that can lower the root-mean-square error (RMSE) of weighted estimates with respect to a specific outcome of interest. Model-assisted calibration provides the framework to construct calibrated weights targeting an outcome variable, given a model that can approximate the expected values of the outcome (Wu and Sitter, 2001). The key to successful model-assisted calibration is a model with strong predictive properties: model parameters estimated from one sample can be used to reliably predict values in a different sample of the same population. Of course, such predictors are not always available; Tourangeau, Conrad and Couper (2013) provide an example where the lack of predictive covariates prevent weighting adjustments from performing well. However, Tourangeau, et al. (2013) had in mind household surveys. Predictors can be more powerful in establishment or institutional surveys or in some specialized surveys like election polls. For example, Wang, Rothschild, Goel and Gelman (2015) use party affiliation and candidate voted in the previous election to make accurate predictions of the outcome of the 2012 US presidential election based on a non-probability sample that was distributed much different from that of all voters.

Clearly, then, model-assisted calibration might be expected to be most effective when there is a relatively rich set of auxiliary population covariates and consequently an extremely large set of models to be considered. In these settings, obtaining balance between structure – to minimize model misspecification and thus bias – and parsimony – to stabilize estimates and thus minimize variance – can be challenging. The

Least Angle Shrinkage and Selection Operator, LASSO, is a regularized regression that can perform both variable selection and parameter estimation (Tibshirani, 1996). A wide range of applications have demonstrated that LASSO is effective in preventing model over-fitting by automatically selecting more accurate and parsimonious models. Kamarianakis, Shen and Wynter (2012) found success with LASSO in predicting average traffic speed in the presence of severe multi-collinearity due to aggregated area-level regressors. Kohannim, Hibar, Stein, Jahanshad, Hua, Rajagopalan, Toga, Jack Jr, Weiner, de Zubicaray and McMahon (2012) applied LASSO regression to identify subsets of high-dimensional and correlated single nucleotide polymorphisms (SNPs) that are related to brain structure measures. In a review of challenges in ecological analysis with collinear covariates, Dormann, Elith, Bacher, Buchmann, Carl, Carre, Marquez, Gruber, Lafourcade, Leitao and Mnkemller (2013) found that LASSO is one of the methods to consistently produce low root-mean-square-errors. In the fields of genetics and finance, LASSO has been used effectively in prediction modeling with hundreds or thousands of predictors (Wu, Chen, Hastie, Sobel and Lange, 2009).

There is a literature that considers stabilizing forms of traditional calibration. Park and Yang (2008) considered a ridge regression form of a generalized regression estimator that used a penalty term to stabilize the calibration estimators, proving design consistency and showing reduction in variance in simulation studies. Goga, Muhammad-Shehzad and Vanheuverzwyn (2011) and Cardot, Goga and Shehzad (2017) considered calibration to principle components of population totals rather than the population totals themselves, allowing large numbers of auxiliary variables to be collapsed into a manageable subset. Perhaps most relevant to this work, McConville (2011) and McConville, Breidt, Lee and Moisen (2017) developed, again under traditional calibration, the theoretical framework to show approximate design unbiasedness and consistency of LASSO calibration estimator of a total, given LASSO regression parameter estimates. Although model-assisted calibration with LASSO holds great promise in constructing a set of weights that can result in small RMSE of weighted estimates for an outcome variable in a non-probability sample, there is no theoretical framework established for the bias and consistency properties of model-assisted LASSO calibration estimators for non-probability sample.

Thus the main objectives of this article are:

(1) Develop the theoretical framework for model-assisted calibration with LASSO for both continuous and binary outcome variables: derive the point estimate of the total, its asymptotic expectation, and asymptotic theoretical variance estimate.

(2) Investigate relative performances, in terms of root-mean-square-error, of LASSO calibration to traditional calibration under different outcome types, sampling schemes, sample sizes, and calibration variable covariance structures.

While our development of the asymptotic theory assumes known design weights, a key finding is that LASSO calibration yields consistent estimators of a population total regardless of whether the design weights are correctly specified as long as the regression model includes all superpopulation parameters as a subset of the parameters in the model. Hence, we focus estimation in the simulation studies in the non-probability-based setting, where initial design weights taken to be the same as those for simple-random-sampling (SRS), $d_i = N/n$ for population and sample sizes N and n, regardless of how the samples are formed (which in practice would be unknown). We also apply LASSO calibration to estimation of the total number of adults diagnosed with cancer in the US population, using data on cancer incidence from the 2013 National Health Interview Survey (NHIS) and auxiliary population data from the US Census American Communities Survey, ignoring sample design weight to approximate a non-probability sample and comparing results to the fully-weighted (representative) estimates.

The organization of this article is as follows. Section 2 provides the definition and notations for calibration and LASSO regression. Section 3 develops the LASSO calibration estimator of population total, its model expectation, and asymptotic variances. Section 4 describes the simulation and results for evaluating the root-mean-square-error and variance estimates of the LASSO-calibrated estimator. Section 5 considers the NHIS example. We conclude with Section 6 summarizing the findings.

2 Calibration

2.1 Traditional calibration

For an analytical sample s_A (the sample which requires weight calibration) of size n drawn from sample design $\mathcal{A}$ with design weights $\underset{n\times 1}{\mathbf{d}}$, and the diagonal matrix of design weights $\mathbf{D}$, calibrated weights $\underset{n\times 1}{\mathbf{w}}$ minimize a distance measure

$$E_{\mathcal{A}}\left[\sum_{i\in s_A} g(w_i, d_i)\Big/ q_i\right] \tag{2.1}$$

under the constraint:

$$\sum_{i\in s_A} w_i \mathbf{x}_i^T = \mathbf{T}^{\mathbf{X}} \tag{2.2}$$

where $E_{\mathcal{A}}$ is expectation with respect to the analytic (probability) design, $g(w_i, d_i)$ is a differentiable function with respect to w_i, strictly convex on an interval containing d_i, and $g(d_i, d_i) = 0$, and where $\mathbf{T}^X$ is a row vector of known population totals of sample calibration variables $\mathbf{X}$ (Deville and Särndal, 1992). The constant q_i is independent of design weight d_i. The commonly used generalized regression (GREG) estimator uses the chi-square distance: $g(w_i, d_i) = (w_i - d_i)^2 \big/ d_i$ with $q_i = 1$. Under this distance measure:

$$\mathbf{w}^{\text{GREG}} = \mathbf{d} + \mathbf{DX}(\mathbf{X}^T\mathbf{DX})^{-1}(\mathbf{T}^{\mathbf{X}} - \mathbf{d}^T\mathbf{X})^T. \tag{2.3}$$

The estimate of population total of outcome $\mathbf{y}$ is based on calibrated weights:

$$\begin{aligned}\hat{T}_y^{\text{GREG}} &= \mathbf{w}^{(\text{GREG})T}\mathbf{y}\\ &= \mathbf{d}^T\mathbf{y} + (\mathbf{T}^{\mathbf{X}} - \mathbf{d}^T\mathbf{X})(\mathbf{X}^T\mathbf{DX})^{-1}\mathbf{X}^T\mathbf{Dy}\\ &= \hat{T}_y^{\text{HT}} + (\mathbf{T}^{\mathbf{X}} - \mathbf{d}^T\mathbf{X})\hat{\boldsymbol{\beta}}\end{aligned} \tag{2.4}$$

where $\hat{T}_y^{\text{HT}} = \sum_{i\in s_A} d_i y_i$ is the standard (weighted) design-based estimator, $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{DX})^{-1}\mathbf{X}^T\mathbf{Dy}$ is the weighted least squares estimate of the linear regression $E_\xi[y_i \mid \mathbf{x}_i, \boldsymbol{\beta}] = \mathbf{x}_i^T\boldsymbol{\beta}$, given weights $\mathbf{D}$. (This corresponds to the poststratified estimator when $\mathbf{X}$ consists entirely of cell totals for categorical variables.)

The calibrated weights defined in equation (2.3) do not rely on any outcome variable. Thus the same set of weights can be applied to all variables in the survey. Note that GREG assumes a working model that is linear. Although $\hat{T}_y^{\text{GREG}}$ is asymptotically design-unbiased for $T_y,$ when the relationship between $\mathbf{y}$ and $\mathbf{X}$ is non-linear, such as in the case when $\mathbf{y}$ is binary, the design variance of $\hat{T}_y^{\text{GREG}}$ can be larger than the design variance $\hat{T}_y^{\text{HT}}.$

2.2 Model-assisted calibration

Model-assisted calibration estimators can have significant advantage over $\hat{T}_y^{\text{GREG}}$ because model-assisted calibration allows for non-linear models to assist in the construction of calibrated weights. In model-assisted calibration, we assume a relationship between an outcome $\mathbf{y}$ and $\mathbf{X}$ through first two moments (Wu and Sitter, 2001):

$$E_\xi\left(y_i \mid \mathbf{x}_i\right) = \mu\left(\mathbf{x}_i, \boldsymbol{\beta}\right), V_\xi\left(y_i \mid \mathbf{x}_i\right) = v_i^2\sigma^2 \tag{2.5}$$

where $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_p\right)^T$ and σ are unknown superpopulation parameters, $\mu\left(x_i, \boldsymbol{\beta}\right)$ is a known function of $\mathbf{x}_i$ and $\boldsymbol{\beta},$ and v_i is a known function of $\mathbf{x}_i$ or $\mu\left(\mathbf{x}_i, \boldsymbol{\beta}\right).$ E_ξ and V_ξ are expectation and variance with respect to the model $\xi.$ Let $\mathbf{B}$ be the finite population (or census) estimate of $\boldsymbol{\beta}$ (i.e., the quasilikelihood estimator of $\boldsymbol{\beta}$ based on the entire finite population), and $\hat{\mu}_i = \mu\left(\mathbf{x}_i, \hat{\mathbf{B}}\right),$ where $\hat{\mathbf{B}}$ is the sample estimate of $\mathbf{B}.$ The model-assisted calibrated weights $\mathbf{w}$ then minimize a distance measure $E_{\mathcal{A}}\left[\sum_{i \in s_A} g\left(w_i, d_i\right)/q_i\right]$ under the constraints $\sum_{i \in s_A} w_i = N$ and $\sum_{i \in s_A} w_i \hat{\mu}_i = \sum_i^N \hat{\mu}_i.$ The main conceptual difference between traditional calibration and model-assisted calibration is that in model-assisted calibration, the constraints are based on two quantities: (1) population size, and (2) population total of predicted values $\hat{\mu}_i.$ In traditional calibration, the constraint is a vector of population totals of $\mathbf{X}$ (see equation (2.2)). Under chi-square distance measure with $q_i = 1,$ the model-assisted calibrated weights are:

$$\mathbf{w}^{\text{MC}} = \mathbf{d} + \mathbf{DM}\left(\mathbf{M}^T\mathbf{DM}\right)^{-1}\left(\mathbf{T}^M - \mathbf{d}^T\mathbf{M}\right)^T \tag{2.6}$$

where $\mathbf{T}^M = \left[N, \sum_i^N \hat{\mu}_i\right]$ and $\mathbf{M} = \left[\mathbf{d}, \left(\hat{\mu}_i\right)_{i \in s_A}\right].$ (In the non-probability setting the vector of design weights $\mathbf{d}$ can be replaced with $(N/n)\,\mathbf{1}.$) The estimate for the population total based on model-assisted calibrated weights is then:

$$\begin{aligned}\hat{T}_y^{\text{MC}} &= \left(\mathbf{w}^{\text{MC}}\right)^T \mathbf{y}\\ &= \mathbf{d}^T\mathbf{y} + \left(\mathbf{T}^{\mathbf{M}} - \mathbf{d}^T\mathbf{M}\right)\left(\mathbf{X}^T\mathbf{DX}\right)^{-1}\mathbf{X}^T\mathbf{Dy}\\ &= \hat{T}_y^{\text{HT}} + \left(\sum_i^N \hat{\mu}_i - \sum_{i \in s_A} d_i\hat{\mu}_i\right)\hat{B}^{\text{MC}}\end{aligned} \tag{2.7}$$

where $\hat{B}^{\text{MC}}$ is the calibration slope to satisfy the calibration constraints (different from the model parameter estimates $\hat{\mathbf{B}}$) :

$$\hat{B}^{\text{MC}} = \frac{\sum_{i \in s_A} d_i\left(\hat{\mu}_i - \hat{\bar{\mu}}\right)\left(y_i - \bar{y}\right)}{\sum_{i \in s_A} d_i\left(\hat{\mu}_i - \hat{\bar{\mu}}\right)^2},\ \hat{\bar{\mu}} = \sum_{i \in s_A} d_i\hat{\mu}_i \Big/ \sum_{i \in s_A} d_i,\ \bar{y} = \sum_{i \in s_A} d_i y_i \Big/ \sum_{i \in s_A} d_i.$$

Unbiasedness and small variances of $\hat{T}_y^{\text{MC}}$ both rely on how well the $\hat{\mu}_i$ approximates the true expected value of $y_i.$

3 Model selection and robust calibration using adaptive LASSO

3.1 Adaptive LASSO background

3.1.1 Definition and parameters

The adaptive LASSO regression coefficients are obtained by solving a penalized regression equation. For linear adaptive LASSO regression (Zou, 2006):

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left(\sum_{i \in s_A} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \lambda_n \sum_{j=1}^{p} \alpha_j^\gamma \,|\, \beta_j \,| \right) \tag{3.1}$$

where α_j^γ is an adjustable weight and λ_n is a penalty used to optimize a model fit measure. Similarly for logistic adaptive LASSO:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left(\sum_{i \in s_A} [-y_i\,(\mathbf{x}_i^T \beta) + \log(1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta}))] + \lambda_n \sum_{j=1}^{p} \alpha_j^\gamma \,|\, \beta_j \,| \right). \tag{3.2}$$

Given λ_n and $\gamma,$ we can calculate $\hat{\boldsymbol{\beta}}$ through iterative procedures. The R package *glmnet* will compute both the linear and logistic adaptive LASSO (Friedman, Hastie and Tibshirani, 2010).

The role of the weight parameter, $\alpha_j,$ is to prevent LASSO from selecting covariates with large effect sizes in favor of lowering prediction error when the sample size is small. Thus the weights are inversely proportional to effect sizes of regression parameters: $\alpha_j \propto 1/|\beta_j|.$ A common choice of α_j is $1/|\hat{\beta}_j^{\text{MLE}}|,$ where $\hat{\beta}_j^{\text{MLE}}$ is the maximum likelihood estimate of $\beta_j.$ The power of the weight parameter, $\gamma,$ is a constant greater than 0 that interacts with α_j to control LASSO from selecting or excluding parameters. For example, if we still want LASSO to favor large effect covariates when the sample size is small, we should set γ small. If we want to de-emphasize effect sizes further, we should set γ large.

3.1.2 Oracle property

An important concept in measuring the performance of a model selection and estimation method is called the "oracle property". The optimal method selects the correct variables and provides unbiased estimates of selected parameters. Suppose the parameters in a full regression model have both zero and non-zero components. Without loss of generality, let the first p be non-zero and the last q zero:

$$\boldsymbol{\beta}^F = \begin{pmatrix} \boldsymbol{\beta}_{(p \times 1)}^{(1)} \\ \boldsymbol{\beta}_{(q \times 1)}^{(2)} = \mathbf{0} \end{pmatrix}.$$

A regression model has the oracle property if it satisfies the following conditions (Fan and Li, 2001):

- The probability of estimating 0 for zero-valued parameters tends to one: $\Pr\left(\hat{\boldsymbol{\beta}}^{(2)} = \mathbf{0}\right) \to 1$ as $n \to \infty.$

- The estimates of non-zero parameters are as good as if the true sub-model is known: $\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(1)}\right) \rightarrow N\left(\mathbf{0}, \mathbf{C}\right)$ where $\mathbf{C} = \Sigma\left(\boldsymbol{\beta}^{(1)}\right)$ is the covariance matrix of $\boldsymbol{\beta}^{(1)}$ under linear model, and $\mathbf{C} = I^{-1}\left(\boldsymbol{\beta}^{(1)}\right)$ is the inverse of the Fisher information matrix of $\boldsymbol{\beta}^{(1)}$ under the generalized linear model.

For finite-population inference, suppose ν indexes a population with size N_ν, let $\mathbf{B}_\nu$ be the quasilikelihood estimates of $\boldsymbol{\beta}$ in population ν, and $\hat{\mathbf{B}}_\nu$ is the estimate of $\mathbf{B}_\nu$ based on a sample with size $n_\nu \leq N_\nu$. We assume that $N_\nu \rightarrow \infty, n_\nu \rightarrow \infty,$ and $n_\nu / N_\nu \rightarrow 0$ as $\nu \rightarrow \infty$. The finite-population equivalent of the oracle property is then:

$$\begin{aligned}
\Pr\left(\hat{\mathbf{B}}_\nu^{(2)} = \mathbf{0}\right) &\rightarrow 1 \\
\sqrt{n_\nu}\left(\hat{\mathbf{B}}_\nu^{(1)} - \mathbf{B}_\nu^{(1)}\right) &\rightarrow N_\nu\left(\mathbf{0}, \mathbf{C}_\nu\right) \\
\mathbf{B}_\nu &\rightarrow \boldsymbol{\beta} \\
\text{as } & \nu \rightarrow \infty
\end{aligned}$$

where $\mathbf{C}_\nu = \Sigma\left(\mathbf{B}_\nu^{(1)}\right)$ is the covariance matrix of $\mathbf{B}_\nu^{(1)}$ if the model is linear, and $\mathbf{C}_\nu = I^{-1}\left(\mathbf{B}_\nu^{(1)}\right)$ is the inverse of Fisher information matrix of $\mathbf{B}_\nu^{(1)}$ under the generalized linear model.

Zou (2006) has shown that if $\lambda_n \Big/ \left(\sqrt{n}\Big/\left(\sqrt{n}\right)^\gamma\right) \rightarrow \infty$ and $\lambda_n \big/ \sqrt{n} \rightarrow 0,$ then the adaptive LASSO satisfies the oracle property. The conditions require that λ_n grow at least at the rate of $\sqrt{n}\Big/\left(\sqrt{n}\right)^\gamma$, but not faster than $\sqrt{n}$. The choice of λ_n and γ, and R code for implementing it, are discussed in the Appendix.

3.2 LASSO calibration

This section derives the analytical formula for a LASSO estimator of total, its model expectation, and estimators of the asymptotic design variance. We make the following assumptions:

1. The samples are drawn from a single-stage sample design $\mathcal{A}$, allowing for unequal probabilities of selection. The selection probability for unit i is denoted by π_i^A, and the joint selection probability of units i and j is denoted by π_{ij}^A. We denote the design weight for unit i by $d_i^A = 1/\pi_i^A$, the vector of design weights by $\mathbf{d}^A$, and the diagonal matrix of design weights by $\mathbf{D}^A$.
2. Population-level auxiliary data are known, denoted by $\mathbf{X} = \left(\mathbf{x}_i^T\right), i = 1, \ldots, N$.
3. A superpopulation model is assumed, as is described in Section 2.2:
$$\begin{aligned}
E_\xi\left(y_i \mid \mathbf{x}_i\right) &= \mu\left(\mathbf{x}_i, \boldsymbol{\beta}\right) \\
V_\xi\left(y_i \mid \mathbf{x}_i\right) &= \nu_i^2 \sigma^2.
\end{aligned}$$
4. The true superpopulation parameters are a subset of the full regression model for LASSO:
$$\boldsymbol{\beta}^F = \begin{pmatrix} \boldsymbol{\beta}_{(p\times 1)} \\ \boldsymbol{\beta}_{(q\times 1)}^{(2)} \end{pmatrix}.$$

5. The full-range of $\mathbf{X}$ in the population has non-zero probability of being observed in the analytical sample.

3.2.1 Point estimate: $\hat{T}_y^{\text{LASSO}}$

The LASSO calibration estimate of total can be obtained following the steps:

1. Obtain LASSO regression coefficients $\hat{\mathbf{B}}$ as described in the Appendix.
2. Use $\hat{\mathbf{B}}$ to calculate $\hat{\mu}_i = \mu\left(\mathbf{x}_i, \hat{\mathbf{B}}\right)$ in the population.
3. Define $\mathbf{T}^M = \left(N, \sum_i^N \hat{\mu}_i\right)$ and $\mathbf{M} = \left[\mathbf{d}^A, \sum_{i \in s_A} \hat{\mu}_i\right]$, under chi-square distance measure with $q_i = 1$:

$$\mathbf{w}^{\text{LASSO}} = \mathbf{d}^A + \mathbf{D}^A \mathbf{M}\left(\mathbf{M}^T \mathbf{D}^A \mathbf{M}\right)^{-1}\left(\mathbf{T}^M - \left(\mathbf{d}^A\right)^T \mathbf{M}\right)^T. \tag{3.3}$$

4. Determine the LASSO calibration estimator of total:

$$\begin{aligned}\hat{T}_y^{\text{LASSO}} &= \left(\mathbf{w}^{\text{LASSO}}\right)^T \mathbf{y}\\ &= \left(\mathbf{d}^A\right)^T \mathbf{y} + \left(\mathbf{T}^{\mathbf{M}} - \left(\mathbf{d}^A\right)^T \mathbf{M}\right)\left(\mathbf{X}^T \mathbf{D}^A \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{D}^A \mathbf{y}\\ &= \left(\mathbf{d}^A\right)^T \mathbf{y} + \left(\sum_i^N \hat{\mu}_i - \sum_{i \in s_A} d_i^A \hat{\mu}_i\right) \hat{B}^{\text{MC}}\end{aligned} \tag{3.4}$$

where $\hat{B}^{\text{MC}}$ is the calibration slope to satisfy the calibration constraints:

$$\hat{B}^{\text{MC}} = \frac{\sum_{i \in s_A} d_i^A \left(\hat{\mu}_i - \hat{\bar{\mu}}\right)\left(y_i - \bar{y}\right)}{\sum_{i \in s_A} d_i^A \left(\hat{\mu}_i - \hat{\bar{\mu}}\right)^2},\ \hat{\bar{\mu}} = \sum_{i \in s_A} d_i^A \hat{\mu}_i \Big/ \sum_{i \in s_A} d_i^A,\ \bar{y} = \sum_{i \in s_A} d_i^A y_i \Big/ \sum_{i \in s_A} d_i^A.$$

3.2.2 Asymptotic behavior of $\hat{T}_y^{\text{LASSO}}$

Wu and Sitter (2001) established the conditions to derive an asymptotic model-assisted calibration estimator. We state the conditions here with slight modification in notations to be consistent with the current research. Let $\boldsymbol{\beta}$ be the true superpopulation parameter for the model defined in equation (2.5), and $\mathbf{B}$ be the finite-population quasilikelihood estimator of $\boldsymbol{\beta}$. The following conditions are used for deriving LASSO calibration asymptotic estimator of total:

1. $\hat{\mathbf{B}} = \mathbf{B} + O_p\left(1/\sqrt{n}\right)$, $\mathbf{B}$ is the finite-population regression slope of $\boldsymbol{\beta}$, $\mathbf{B} \to \boldsymbol{\beta}$.
2. For each $\mathbf{x}_i$, $\partial\mu(\mathbf{x}_i, \mathbf{t})/\partial\mathbf{t}$ is continuous in $\mathbf{t}$, and $\max_i \left|\partial\mu(\mathbf{x}_i, \mathbf{t})/\partial\mathbf{t}\right| \le h(\mathbf{x}_i, \boldsymbol{\beta})$ for $\mathbf{t}$ in a neighborhood of $\boldsymbol{\beta}$, and $N^{-1}\sum_{i \in U} h(\mathbf{x}_i, \boldsymbol{\beta}) = O(1)$.
3. For each $\mathbf{x}_i$, $\partial^2\mu(\mathbf{x}_i, \mathbf{t})/\partial\mathbf{t}\partial\mathbf{t}^T$ is continuous in $\mathbf{t}$, and $\max_{j,k} \left|\partial^2\mu(\mathbf{x}_i, \mathbf{t})/\partial t_j \partial t_k\right| \le k(\mathbf{x}_i, \boldsymbol{\beta})$ for $\mathbf{t}$ in a neighborhood of $\boldsymbol{\beta}$, and $N^{-1}\sum_{i \in U} k(\mathbf{x}_i, \boldsymbol{\beta}) = O(1)$.
4. The Horvitz-Thompson (HT) estimators of certain population means are asymptotically normally distributed (Fuller, 2009; pages 47-57).
5. $\lambda_n \Big/ \left(\sqrt{n} \big/ \left(\sqrt{n}\right)^{\gamma}\right) \to \infty$ and $\lambda_n / \sqrt{n} \to 0$.

Lemma 1: Assume that superpopulation model (2.5) holds. Let $\mathbf{B}$ be the finite-population quasilikelihood estimate of $\boldsymbol{\beta}$, $\mathbf{B} \to \boldsymbol{\beta}$. Under conditions (1)-(5), the model-assisted asymptotic estimator of population total is:

$$\hat{T}_y^{\mathrm{MC}} = \sum_{i \in s_A} d_i^A \left(y_i - \mu_i B^{\mathrm{MC}} \right) + \sum_{i=1}^{N} \mu_i B^{\mathrm{MC}} + o_p \left(\frac{N}{\sqrt{n}} \right) \tag{3.5}$$

where

$$\begin{aligned} \mu_i &= \mu(\mathbf{x}_i, \mathbf{B}) \\ B^{\mathrm{MC}} &= \frac{\sum_{i=1}^{N} (\mu_i - \bar{\mu})(y_i - \bar{y})}{\sum_{i=1}^{N} (\mu_i - \bar{\mu})^2}. \end{aligned}$$

Proof. See Appendix.

Given Lemma 1, we derive $\hat{T}_y^{\mathrm{LASSO}}$ the asymptotic LASSO estimator of total in Theorem 1. We show $\hat{T}_y^{\mathrm{LASSO}}$ is model unbiased for the population total in Theorem 2. Finally, Theorem 3 determines variance estimates for the LASSO calibration estimator of a total.

Theorem 1: Suppose the parameters in a full regression model have both zero and non-zero components. Without loss of generality, let the first p be non-zero and the last q be zero:

$$\boldsymbol{\beta}^F = \begin{pmatrix} \boldsymbol{\beta}_{(p\times 1)}^{(1)} \\ \boldsymbol{\beta}_{(q\times 1)}^{(2)} \end{pmatrix}, \quad \boldsymbol{\beta}^{(1)} = \boldsymbol{\beta} \quad \text{and} \quad \boldsymbol{\beta}^{(2)} = \mathbf{0}_{(q\times 1)},$$

under conditions (1)-(5), the asymptotic LASSO calibration estimator of total is:

$$\hat{T}_y^{\mathrm{LASSO}} = \sum\nolimits_{i \in s_A} d_i^A \left(y_i - \mu_i B^{\mathrm{MC}} \right) + \sum\nolimits_{i=1}^{N} \mu_i B^{\mathrm{MC}} + o_p \left(\frac{N}{\sqrt{n}} \right).$$

Proof. See Appendix.

Theorem 2: $\hat{T}_y^{\mathrm{LASSO}}$ is model-unbiased, that is $E_\xi \left(\hat{T}_y^{\mathrm{LASSO}} \right) = T$.

Proof. See Appendix.

Thus, as long as LASSO regression parameters include the superpopulation parameters, $\hat{T}_y^{\mathrm{LASSO}}$ is model-unbiased regardless of design weights. (Note that this is a quality that $\hat{T}_y^{\mathrm{GREG}}$ shares with $\hat{T}_y^{\mathrm{LASSO}}$. However, $\hat{T}_y^{\mathrm{LASSO}}$ can assume models with much larger numbers of covariates than $\hat{T}_y^{\mathrm{GREG}}$.) This property is essential in non-probability samples, where there are no initial design weights to guarantee unbiasedness.

Theorem 3: The estimator of the asymptotic variance of $\hat{T}_y^{\mathrm{LASSO}}$ is given by

$$\begin{aligned} v_{\mathcal{A}} \left(\hat{T}_y^{\mathrm{LASSO}} \right) &= \sum_{i \in s_A} \left(\frac{y_i - \hat{\mu}_i \hat{B}^{\mathrm{MC}}}{\pi_i} \right)^2 (1 - \pi_i) \\ &\quad + \sum_{i \in s_A} \sum_{j \neq i} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{\left(y_i - \hat{\mu}_i \hat{B}^{\mathrm{MC}} \right)}{\pi_i} \frac{\left(y_j - \hat{\mu}_j \hat{B}^{\mathrm{MC}} \right)}{\pi_j}. \end{aligned} \tag{3.6}$$

Proof. The theoretical design variance of the LASSO estimator is

$$\begin{aligned} V_{\mathcal{A}}\left(\hat{T}_y^{\text{LASSO}}\right) &= V_{\mathcal{A}}\left(\sum_{i\in s_A} d_i^A\left(y_i - \mu_i B^{\text{MC}}\right) + \sum_{i=1}^{N} \mu_i B^{\text{MC}}\right) \\ &= V_{\mathcal{A}}\left(\sum_{i\in s_A} d_i^A\left(y_i - \mu_i B^{\text{MC}}\right)\right) \\ &= \sum_{i\in U}\left(\frac{y_i - \mu_i B^{\text{MC}}}{\pi_i}\right)^2 \pi_i\left(1-\pi_i\right) \\ &\quad + \sum_{i\in U}\sum_{j\neq i}\left(\pi_{ij} - \pi_i\pi_j\right)\frac{\left(y_i - \mu_i B^{\text{MC}}\right)}{\pi_i}\frac{\left(y_j - \mu_j B^{\text{MC}}\right)}{\pi_j} \end{aligned} \tag{3.7}$$

which follows from equation (3.30) derived for the variance of traditional LASSO calibration estimator of total in McConville (2011). Equation (3.6) then follows from replacing estimates for population quantities.

An alternative variance estimate, suggested by Särndal, Swensson and Wretman (1989), multiplies $\left(y_i - \hat{\mu}_i\hat{B}^{\text{MC}}\right)$ by g – weights, which are the ratios of calibrated weights to the original design weights:

$$\mathbf{g} = \mathbf{1}_{(n\times 1)} + \mathbf{M}\left(\mathbf{M}^T\mathbf{D}^A\mathbf{M}\right)^{-1}\left(\mathbf{T}^M - \left(\mathbf{d}^A\right)^T\mathbf{M}\right)^T$$

$$\begin{aligned} v.g_{\mathcal{A}}\left(\hat{T}_y^{\text{LASSO}}\right) &= \sum_{i\in s_A}\left(\frac{g_i\left(y_i - \hat{\mu}_i\hat{B}^{\text{MC}}\right)}{\pi_i}\right)^2\left(1-\pi_i\right) \\ &\quad + \sum_{i\in s_A}\sum_{j\neq i}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}}\frac{g_i\left(y_i - \hat{\mu}_i\hat{B}^{\text{MC}}\right)}{\pi_i}\frac{g_j\left(y_j - \hat{\mu}_j\hat{B}^{\text{MC}}\right)}{\pi_j}. \end{aligned} \tag{3.8}$$

To simplify notations, we refer to $v_{\mathcal{A}}\left(\hat{T}_y^{\text{LASSO}}\right)$ as v^{LASSO} and $v.g_{\mathcal{A}}\left(\hat{T}_y^{\text{LASSO}}\right)$ as v_g^{LASSO}.

4 Simulation study

We design a simulation to evaluate the finite sample properties of $\hat{T}_y^{\text{LASSO}}$ and the asymptotic variance estimates of $\hat{T}_y^{\text{LASSO}}$, v^{LASSO} and v_g^{LASSO}. We also consider a naive bootstrap estimator $v_{\text{boot}}^{\text{LASSO}}$, obtained by drawing 500 samples with replacement from each simulation sample, as an alternative variance estimator of $\hat{T}_y^{\text{LASSO}}$.

To simulate non-probability samples, we generate samples with unequal selection probabilities, but set design weights to $\mathbf{d}^A = N/n$. We also consider $\hat{T}_y^{\text{GREG}}$ (traditional calibration estimator) and $\hat{T}_y^{\text{HT}}$ (pure design-based Horvitz-Thompson estimator). Because $\hat{T}_y^{\text{LASSO}}$ performs both variable selection and estimation, we implement a backward stepwise selection to select the working model for GREG. Although there is no theoretical justification for using stepwise variable selection, Skinner and Silva (1997) have shown that given two auxiliary variables, a stepwise procedure can result in improved efficiency of GREG estimator. We are interested in knowing the performance of each estimator under (1) populations with different signal-to-noise-ratios (SNR), (2) independent, informative, and biased sampling schemes, and (3) small and large sample sizes. The signal-to-noise ratio is calculated according to definitions in Czanner, Sarma, Eden and Brown (2008). We set two levels of correlations (low/high) between covariates, crossed with two levels of effect sizes (low/high) of the covariates. We set the low/high and high/low populations to have the same SNR in order to understand the influence of correlation and effect size on estimator's performance given the same SNR. Three sampling schemes are used to draw samples: simple-random-sampling without replacement, SRS, Poisson sampling with selection probabilities proportional to covariates, $\mathrm{POI}(\mathrm{X})$, and Poisson sampling with selection probabilities proportional to covariates and the outcome, $\mathrm{POI}(\mathrm{X+Y})$. $\mathrm{POI}(\mathrm{X+Y})$ sampling simulates self-selection bias of non-probability samples, where the propensity of a respondent to participate in a study relates to the analysis variable. We consider two sample sizes: 250 and 1,000. Thus we have a total of $2 \times 2 \times 3 \times 2 = 24$ experimental groups.

4.1 Population

To create collinearity among covariates, we follow an auto-decay correlation structure commonly used in LASSO-related simulations (Tibshirani, 1996): $\operatorname{cor}(X_i, X_j) = \rho^{|i-j|}$, $i = 1, \ldots, p$. We generate a population of size $N = 100{,}000$ from a multivariate normal distribution with mean $\mathbf{0}_{(p\times 1)}$ and covariance Σ^{ρ}, $p = 40$. The continuous outcome variable is generated by the regression model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{40} x_{i40} + N(0,3).$$

The binary outcome variable is generated by the logistic regression model:

$$\begin{aligned} \phi_i &= \operatorname{expit}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{40} x_{i40}), \quad \operatorname{expit}(u) = (1+\exp(u))^{-1} \\ y_i &= \operatorname{bernoulli}(\phi_i). \end{aligned}$$

We set $\rho = 0.15$ for low correlation population, and $\rho = 0.73$ for high correlation population. For both continuous and binary outcome variables:

$$\begin{aligned} \text{Low effect-size } \boldsymbol{\beta}^{(1)} &:= \beta_{12} \ldots \beta_{19}, \beta_{32} \ldots \beta_{39} = 0.45 \\ \text{High effect-size } \boldsymbol{\beta}^{(1)} &:= \beta_{12} \ldots \beta_{19}, \beta_{32} \ldots \beta_{39} = 0.74. \end{aligned}$$

For continuous $\mathbf{y}$: $\beta_0 = 1$, for binary $\mathbf{y}$: $\beta_0 = 0.4$. The rest of $\beta_i = 0$. Out of 41 regression parameters, 16 are non-zero and 25 are zero.

4.2 Sampling schemes

Three sampling schemes are used to generate the sample:

1. Simple-Random-Sampling (SRS): selection probabilities $= n/N$.
2. Poisson sampling with probabilities proportional to $\mathbf{X}$, $\mathrm{POI}(\mathrm{X})$

$$\begin{cases} \text{continuous } \mathbf{y}: \ \pi_i \propto 0.4 + 0.4x_{i5} + 0.4x_{i15} + 0.4x_{i25} + 0.4x_{i35} \\ \quad \text{binary } \mathbf{y}: \ \operatorname{logit}(\pi_i) = 0.4 + 0.4x_{i5} + 0.4x_{i15} + 0.4x_{i25} + 0.4x_{i35}. \end{cases}$$

3. Poisson sampling with probabilities proportional to $\mathbf{X}$ and $\mathbf{y}$, $\mathrm{POI}(\mathrm{X+Y})$

$$\begin{cases} \text{continuous } \mathbf{y}: \ \pi_i \propto 0.4 + 0.4x_{i5} + 0.4x_{i15} + 0.4x_{i25} + 0.4x_{i35} + 0.5y_i \\ \quad \text{binary } \mathbf{y}: \ \mathrm{logit}(\pi_i) \propto 0.4 + 0.4x_{i5} + 0.4x_{i15} + 0.4x_{i25} + 0.4x_{i35} + y_i. \end{cases}$$

4.3 Evaluation metrics

We evaluate empirical bias, variance, and RMSE for each estimator of total. We evaluate the asymptotic variance estimates and bootstrap variance estimates by their 95% nominal coverage and %bias relative to empirical variance. We use the normal approximation to generate confidence intervals. We calculate %bias as $\%\mathrm{bias} = 100[v - \mathrm{var}(\hat{T}_y^{\mathrm{LASSO}})] \big/ \mathrm{var}(\hat{T}_y^{\mathrm{LASSO}})$, where $\mathrm{var}(\hat{T}_y^{\mathrm{LASSO}})$ is the empirical variance obtained from the simulation samples.

4.4 Simulation results

The simulation results are based on S = 1,000 simulated samples per each experimental group. Table 4.1 lists the numerical results of bias, variance, and root-mean-square-error of each estimator under different experimental designs for estimating the total of a continuous outcome variable. Table 4.2 lists the numerical results for estimating the total of a binary outcome variable.

4.4.1 Root mean square error

Under SRS, all estimators are unbiased, and LASSO and GREG perform approximately equally well relative to HT. $\mathrm{POI}(\mathrm{X})$ and $\mathrm{POI}(\mathrm{X+Y})$ induce biased samples by selecting cases with larger covariate values with higher probabilities. Under $\mathrm{POI}(\mathrm{X+Y})$, the selection also favors cases with larger outcome values. The absolute bias of LASSO decreases relative to GREG as SNR increases. This improvement is more dramatic in the binary case than the continuous case, especially for $\mathrm{POI}(\mathrm{X+Y})$. In terms of RMSE, LASSO has marginal improvement over GREG for estimating totals of continuous outcome variables. The improvement is slightly noticeable, about 3%, when there are highly correlated predictors in the model. For the binary setting, there is substantial improvement in MSE for LASSO over GREG as SNR increases, with reductions of 20% for the $\mathrm{POI}(\mathrm{X})$ and nearly 50% for the $\mathrm{POI}(\mathrm{X+Y})$ setting when SNR is large. In particular, under Low/High and High/Low population types, the SNR is the same, thus the difference in performance between LASSO and GREG is attributed to correlation or effect size. LASSO performs better in both bias and RMSE in High/Low population type, suggesting that LASSO has stronger advantage over GREG when there are highly correlated predictors in the model. This suggests that LASSO has a better variable selection capability in the presence of multicollinearity relative to stepwise variable selection procedure used in GREG.

Table 4.1
Simulation summary for continuous outcome: total, bias, and RMSE $\times 10^3$; variance $\times 10^6$

Population	n	Sampling scheme	HT			GREG			LASSO		
			bias	var	rmse	bias	var	rmse	bias	var	rmse
low/low T = 100.8 SNR = 0.47	250	SRS	0.5	546	23.3	0.9	425	20.6	0.9	428	20.7
		POI(X)	12.4	525	26.0	-0.6	446	21.1	-0.4	441	21.0
		POI(X+Y)	19.4	519	29.9	4.6	443	21.5	4.7	431	21.3
	1,000	SRS	0.2	129	11.4	0.3	94	9.6	0.3	94	9.7
		POI(X)	12.6	129	17.0	-0.1	91	9.5	-0.2	92	9.6
		POI(X+Y)	19.7	128	22.7	4.9	91	10.7	5.0	91	10.7
low/high T = 101.4 SNR = 1.26	250	SRS	0.4	849	29.1	0.9	415	20.4	1.0	417	20.4
		POI(X)	21.1	818	35.6	-1.3	434	20.9	-1.0	432	20.8
		POI(X+Y)	31.7	817	42.7	3.7	427	21.0	4.0	427	21.1
	1,000	SRS	0.0	200	14.1	0.3	94	10.0	0.3	93	9.7
		POI(X)	21.1	199	25.4	-0.1	91	9.6	-0.2	90	9.6
		POI(X+Y)	31.7	196	34.6	4.9	91	10.7	4.8	89	10.6
high/low T = 101.8 SNR = 1.26	250	SRS	0.1	941	30.7	1.0	421	20.6	1.0	399	20.0
		POI(X)	50.2	895	58.5	-0.7	434	20.8	-1.6	402	20.1
		POI(X+Y)	57.8	872	64.9	4.0	435	21.2	3.0	399	20.2
	1,000	SRS	0.0	218	14.8	0.3	94	9.7	0.3	93	9.6
		POI(X)	50.6	210	53.0	-0.1	93	9.7	-0.5	91	9.6
		POI(X+Y)	58.2	209	59.9	4.7	95	10.8	4.2	92	10.5
high/high T = 103.1 SNR = 3.41	250	SRS	-0.4	1,897	43.6	0.8	436	20.9	1.0	407	20.2
		POI(X)	83.3	1,826	93.7	-0.8	435	20.9	-1.5	406	20.2
		POI(X+Y)	96.4	1,779	105.3	3.7	428	21.0	3.0	404	20.3
	1,000	SRS	-0.2	444	21.0	0.3	93	9.7	0.3	93	9.7
		POI(X)	83.6	424	86.1	-0.2	93	9.7	-0.5	91	9.6
		POI(X+Y)	96.9	423	99.0	4.4	94	10.6	4.1	92	10.4

Table 4.2
Simulation summary for binary outcome: total, bias, and RMSE $\times 10^3$; variance $\times 10^6$

Population	n	Sampling scheme	HT			GREG			LASSO		
			bias	var	rmse	bias	var	rmse	bias	var	rmse
low/low T = 56.2 SNR = 0.51	250	SRS	0.0	10.2	3.2	0.0	7.2	2.7	0.0	7.0	2.7
		POI(X)	2.6	10.0	4.1	0.2	8.0	2.8	0.1	7.8	2.8
		POI(X+Y)	4.9	9.8	5.8	2.0	8.1	3.5	1.8	7.8	3.3
	1,000	SRS	-0.0	2.7	1.6	0.0	1.7	1.3	0.0	1.6	1.3
		POI(X)	2.5	2.4	2.9	0.0	1.8	1.3	-0.0	1.7	1.3
		POI(X+Y)	4.7	2.3	5.0	1.8	1.8	2.2	1.6	1.7	2.1
low/high T = 54.4 SNR = 1.10	250	SRS	-0.0	10.8	3.3	0.0	6.1	2.5	0.1	5.4	2.3
		POI(X)	3.0	10.2	4.4	0.1	6.1	2.5	0.1	5.8	2.4
		POI(X+Y)	5.3	9.8	6.2	1.6	6.2	2.9	1.3	5.8	2.8
	1,000	SRS	-0.0	2.7	1.6	0.0	1.3	1.1	0.0	1.1	1.0
		POI(X)	2.9	2.4	3.3	0.0	1.4	1.2	-0.1	1.2	1.1
		POI(X+Y)	5.2	2.2	5.4	1.4	1.4	1.8	1.1	1.2	1.6
high/low T = 54.2 SNR = 1.10	250	SRS	-0.0	10.3	3.2	0.0	5.8	2.4	0.1	4.9	2.2
		POI(X)	6.6	9.6	7.3	0.3	6.2	2.5	-0.2	4.8	2.2
		POI(X+Y)	8.6	9.3	9.1	1.8	6.3	3.1	0.9	4.9	2.4
	1,000	SRS	-0.0	2.5	1.6	0.0	1.2	1.1	0.0	1.0	1.0
		POI(X)	6.6	2.2	6.7	0.2	1.4	1.2	-0.2	1.1	1.1
		POI(X+Y)	8.5	2.1	8.7	1.6	1.4	2.0	1.0	1.0	1.4
high/high T = 52.8 SNR = 2.75	250	SRS	-0.1	10.2	3.1	-0.0	5.2	2.3	0.1	3.8	1.9
		POI(X)	7.1	9.8	7.8	0.3	5.7	2.4	-0.2	3.6	1.9
		POI(X+Y)	9.1	9.4	9.6	1.5	5.7	2.8	0.5	3.7	2.0
	1,000	SRS	-0.1	2.5	1.6	-0.0	1.1	1.0	0.0	0.6	0.8
		POI(X)	7.1	2.2	7.2	0.2	1.3	1.1	-0.2	0.7	0.9
		POI(X+Y)	9.1	2.2	9.2	1.4	1.2	1.8	0.5	0.7	1.0

4.4.2 LASSO variance estimates

Tables 4.3 and 4.4 list the 95% nominal coverage and percent-bias for each of the two asymptotic closed-form variance estimators developed in this research, as well as the naive bootstrap variance estimate of the LASSO calibration estimator.

For continuous outcomes, bootstrap variances have coverages that are consistently close to 95% under SRS and $\text{POI}(\text{X})$ sampling schemes for both sample sizes. Under $\text{POI}(\text{X+Y})$ sampling scheme, there is very modest undercoverage in Table 4.3. The closed-form variances have coverages that are sensitive to both sample size and sampling scheme, with smaller samples tending to undercover, particularly for the $\text{POI}(\text{X+Y})$ sampling scheme. The difference in coverage of variance estimates between small and large sample sizes is expected, since the variance estimates are asymptotic and improve over larger samples. In terms of bias of variance estimators, there is evidence that bias reduces as SNR increases. With the same SNR, both asymptotic closed-form and bootstrap variances have smaller bias given predictors with high correlations relative to predictors with high effect sizes. Closed-form variances tend to underestimate the empirical variance, especially when the sample size is small. Overall, there is very little difference between the two closed-form variance estimates. Bootstrap variance tends to overestimate the empirical variance, but the absolute bias is generally smaller than those of the closed-form variance estimates.

For binary outcomes, both asymptotic closed-form and bootstrap variance estimates are sensitive to sample size, sampling scheme, and SNR. Bootstrap variance coverages are consistently close to 95% under SRS and $\text{POI}(\text{X})$ for both sample sizes and all population types, but coverages range from 75% to 94% under $\text{POI}(\text{X+Y})$. Under $\text{POI}(\text{X+Y})$, the bootstrap variance coverages are better with sample size 250 than with sample size 1,000 when the bias becomes a larger part of the RMSE, and better with high-correlation populations than with low-correlation populations. In terms of coverage, closed-form variances show a similar trend under $\text{POI}(\text{X+Y})$ as bootstrap: better coverage with smaller samples than bigger samples, and better coverage with high-correlation populations than with low-correlation populations. Under SRS and $\text{POI}(\text{X})$, closed-form variance coverage improves as sample size increases. In terms of bias, both bootstrap and closed-form variances have smaller bias with larger sample sizes. Holding sample size fixed, closed-form variance estimates have larger bias as SNR increases. The same trend is not observed in bootstrap variance estimates. Similar to continuous outcome results, closed-form variance tends to underestimate the empirical variance, especially when the sample size is small. Unlike continuous outcome results, there is evidence that the $g-$weighted closed-form variance estimates have better bias-properties than unweighted closed-form variance estimates. The bootstrap variance tends to overestimate the empirical variance. However, the biases are much smaller than for the closed-form variance estimates.

Table 4.3
95% nominal coverage and %bias of variance estimates for LASSO

Continuous outcome			**coverage**			**%bias**		
Population	**n**	**scheme**	v^{LASSO}	v_g^{LASSO}	$v_{\text{boot}}^{\text{LASSO}}$	v^{LASSO}	v_g^{LASSO}	$v_{\text{boot}}^{\text{LASSO}}$
low/low	250	SRS	91.7%	91.8%	95.4%	-22.6%	-22.3%	2.9%
		POI(X)	91.2%	91.2%	96.1%	-25.1%	-24.5%	5.7%
		POI(X+Y)	89.6%	89.9%	95.4%	-23.5%	-22.8%	7.9%
	1,000	SRS	93.2%	93.2%	93.8%	-7.3%	-7.2%	-0.3%
		POI(X)	94.0%	93.9%	95.5%	-5.7%	-5.3%	6.6%
		POI(X+Y)	90.0%	90.1%	92.1%	-4.9%	-4.4%	7.9%
low/high	250	SRS	91.5%	91.5%	95.7%	-22.6%	-22.3%	6.2%
		POI(X)	90.9%	91.2%	96.4%	-25.4%	-24.9%	8.8%
		POI(X+Y)	90.0%	90.2%	95.1%	-24.5%	-23.7%	9.9%
	1,000	SRS	93.4%	93.5%	94.3%	-6.6%	-6.5%	-0.1%
		POI(X)	94.1%	94.2%	95.9%	-4.0%	-3.5%	7.6%
		POI(X+Y)	90.7%	90.7%	92.7%	-2.9%	-2.3%	9.6%
high/low	250	SRS	92.3%	92.2%	95.4%	-17.4%	-17.1%	2.0%
		POI(X)	92.5%	92.6%	95.8%	-17.9%	-16.1%	6.4%
		POI(X+Y)	91.2%	91.8%	96.5%	-17.4%	-15.4%	7.1%
	1,000	SRS	93.5%	93.5%	94.4%	-6.5%	-6.4%	-0.9%
		POI(X)	94.1%	94.0%	95.4%	-5.0%	-3.1%	5.7%
		POI(X+Y)	91.9%	92.3%	93.4%	-6.0%	-3.9%	5.0%
high/high	250	SRS	92.3%	92.3%	95.2%	-19.6%	-19.3%	2.2%
		POI(X)	92.0%	92.3%	96.1%	-19.6%	-17.8%	7.4%
		POI(X+Y)	91.2%	91.8%	95.6%	-19.1%	-16.9%	8.3%
	1,000	SRS	93.4%	93.4%	94.5%	-6.5%	-6.4%	-0.7%
		POI(X)	94.0%	94.5%	95.6%	-4.7%	-2.8%	6.7%
		POI(X+Y)	92.2%	92.4%	93.4%	-5.6%	-3.3%	6.1%

Table 4.4
95% nominal coverage and %bias of variance estimates for LASSO

Binary outcome			**coverage**			**%bias**		
Population	**n**	**scheme**	v^{LASSO}	v_g^{LASSO}	$v_{\text{boot}}^{\text{LASSO}}$	v^{LASSO}	v_g^{LASSO}	$v_{\text{boot}}^{\text{LASSO}}$
low/low	250	SRS	89.8%	90.0%	95.9%	-28.1%	-27.8%	9.2%
		POI(X)	88.1%	88.6%	96.7%	-37.3%	-35.3%	9.2%
		POI(X+Y)	79.0%	79.9%	91.2%	-38.7%	-35.9%	8.0%
	1,000	SRS	92.8%	92.8%	93.5%	-11.9%	-11.8%	-3.5%
		POI(X)	92.0%	92.8%	95.7%	-17.9%	-15.5%	1.0%
		POI(X+Y)	68.6%	69.6%	74.6%	-18.5%	-14.9%	0.5%
low/high	250	SRS	86.8%	87.0%	94.9%	-37.7%	-37.3%	11.3%
		POI(X)	85.4%	86.1%	95.5%	-42.9%	-41.2%	14.4%
		POI(X+Y)	78.7%	80.1%	92.6%	-44.0%	-41.3%	14.4%
	1,000	SRS	94.4%	94.3%	95.2%	-5.5%	-5.4%	5.8%
		POI(X)	91.8%	92.1%	94.9%	-20.5%	-18.6%	-1.8%
		POI(X+Y)	76.8%	77.8%	82.9%	-20.4%	-16.9%	-1.3%
high/low	250	SRS	89.2%	89.1%	94.4%	-28.5%	-28.1%	0.4%
		POI(X)	89.0%	90.1%	95.5%	-31.9%	-25.3%	12.7%
		POI(X+Y)	85.7%	88.4%	93.8%	-33.9%	-25.4%	10.9%
	1,000	SRS	93.9%	93.9%	95.6%	-6.3%	-6.2%	3.5%
		POI(X)	92.6%	93.4%	94.8%	-16.5%	-9.2%	1.9%
		POI(X+Y)	83.3%	85.4%	88.1%	-15.0%	-5.0%	5.2%
high/high	250	SRS	82.8%	82.8%	93.8%	-44.6%	-44.3%	-6.4%
		POI(X)	83.6%	85.5%	95.1%	-44.3%	-39.4%	3.8%
		POI(X+Y)	82.9%	85.1%	93.8%	-45.1%	-38.4%	4.6%
	1,000	SRS	94.3%	94.4%	96.1%	-7.8%	-7.6%	6.3%
		POI(X)	91.3%	92.2%	94.0%	-20.0%	-13.8%	0.2%
		POI(X+Y)	86.3%	88.6%	91.5%	-18.1%	-9.2%	2.8%

5 Application to National Health Interview Survey (NHIS)

5.1 NHIS and ACS data

We next apply LASSO calibration to National Health Interview Survey (NHIS) 2013 to estimate the total number of adults (age 18 or older) diagnosed with cancer in the population. The National Health Interview Survey is a nationally representative sample of non-institutionalized civilian households collected by a multi-stage area-probability sampling (Centers for Disease Control and Prevention, 2005). Each month, health-related data on a cross-sectional sample of people in selected households are obtained by face-to-face interviews. The data provides pseudo-primary-sampling-unit (PSU), pseudo-strata, and sampling weights to allow for weighted estimates with complex survey design. In addition to health-related measures, NHIS also collects demographic data. Our goal is to assess our LASSO estimator by treating the unweighted NHIS sample as reflective of a non-probability sample, and explore how GREG and LASSO calibration compare with the design-weighted estimator.

To calibrate NHIS on a set of demographic and income-related variables, we use the American Community Survey (ACS) 2013 micro-data as the benchmark data. ACS samples are households selected through multi-stage area-probability sampling from 3,143 counties of the U.S. The design of ACS is to improve estimates of small areas between the decennial census long-form samples. Around three million households are selected each year, with measures collected on household types and individual demographics within the households. ACS also collects data from group-quarters, which are excluded from this analysis. For ACS 2013, the sample size for adults is 2,317,301. The NHIS 2013 sample size is 34,201 after removing 242 cases with missing values on demographic variables. For the purposes of this analysis, we treat the weighted estimates from the ACS as known population totals, a reasonable assumption given the differences in sample size.

5.2 Estimators

The outcome variable of interest is whether a person has been diagnosed with cancer. Define the binary indicator for the outcome variable:

$$y_i = \begin{cases} 1: & \text{if person } i \text{ has been diagnosed with cancer} \\ 0: & \text{otherwise.} \end{cases}$$

We first use the NHIS 2013 sampling weights, $\mathbf{w}^{\text{NHIS}}$, and design variables to obtain an unbiased estimate of the population total, $T_y = \sum_{i=1}^{N} y_i$. Then we assume that the NHIS 2013 sample is collected from a simple-random-sampling, with initial design weights $\mathbf{d}^A = N/n$, where N is the population total obtained from ACS, and n is the sample size of NHIS. We calibrate $\mathbf{d}^A$ by a set of demographic and income variables with traditional GREG calibration and LASSO calibration. Finally, as a compromise between GREG and LASSO, we consider model-assisted calibration to a linear model for y_i instead of the LASSO using (2.7);

note that, when $\hat{\mu}_i$ is computed using the same linear model as in GREG, the point estimates of the total will correspond, even though the calibration weights will differ. Thus, we generate seven estimates:

1. $\hat{T}_y^{\text{NHIS}} = \sum_{i \in s_A} w_i^{\text{NHIS}} y_i$: Estimate obtained with NHIS weights.
2. $\hat{T}_y^{\text{HTSRS}} = \sum_{i \in s_A} (N/n)\, y_i$: Estimate obtained with weights $\mathbf{d}^A = N/n$.
3. $\hat{T}_y^{\text{GREG1}} = \sum_{i \in s_A} w_i^{\text{GREG1}} y_i$: Estimate obtained by calibrating $\mathbf{d}^A$ with GREG using all calibration variables.
4. $\hat{t}_y^{\text{GREG1MC}} = \sum_{i \in s_A} w_i^{\text{GREG1MC}} y_i$: Estimate obtained by model-assisted calibration to linear model using predictors in GREG1.
5. $\hat{T}_y^{\text{GREG2}} = \sum_{i \in s_A} w_i^{\text{GREG2}} y_i$: Estimate obtained by calibrating $\mathbf{d}^A$ with GREG using only calibration variables chosen using backward stepwise variable selection.
6. $\hat{t}_y^{\text{GREG2MC}} = \sum_{i \in s_A} w_i^{\text{GREG2MC}} y_i$: Estimate obtained by model-assisted calibration to linear model using predictors in GREG2.
7. $\hat{T}_y^{\text{LASSO}} = \sum_{i \in s_A} w_i^{\text{LASSO}} y_i$: Estimate obtained by model-assisted calibration with LASSO.

The variance of $\hat{T}_y^{\text{NHIS}}$ is the linearization variance estimate of total, accounting for sampling-stratum, primary-sampling-units, and survey weights in the NHIS 2013 sample. Variances of HTSRS, GREG1, and GREG2 are linearization variance estimates with weights $\mathbf{d}^A$, $\mathbf{w}^{\text{GREG1}}$, and $\mathbf{w}^{\text{GREG2}}$ respectively. We obtain the variance of LASSO estimator through naive bootstrap.

5.3 Working models

Table 5.1 lists calibration variable names, labels, values, and distributions in this analysis. The first column is the unweighted distribution of variables in the NHIS sample. The second column contains variable distributions in the NHIS sample, weighted by $\mathbf{w}^{\text{NHIS}}$ person-level weights. The third column is the distribution of variables in the population obtained from the ACS benchmark data. Missing income category is included as a separate category to capture the difference in missing patterns between NHIS and ACS. Including a missing category also allows us to maintain the analytic sample size. Relative to ACS, the unweighted NHIS sample has higher proportions of females, widowed/divorced/separated, and fewer proportion of non-Hispanic whites. After weighting, the NHIS distributions of gender and race are close to the benchmark’s, and only marital status categories show some differences.

We use an unweighted linear model with backward-stepwise variable selection to determine the working model for GREG2. The final variables included in the model for GREG2 are age, education, race, employment status (yes/no), and family income. For standard GREG and LASSO calibration, we use all available variables.

Table 5.1
Calibration variables

		No weights	NHIS Person-level weights	ACS Person-level weights
Region	Northeast	16%	18%	18%
	Midwest	20%	23%	21%
	South	37%	37%	37%
	West	26%	23%	23%
Age	18-29	19%	21%	21%
	30-39	17%	17%	17%
	40-49	16%	18%	18%
	50-59	17%	18%	18%
	60-69	15%	14%	14%
	70-79	9%	8%	8%
	80+	6%	4%	5%
Gender	Male	45%	48%	48%
	Female	55%	52%	52%
Education	Less than high school	16%	14%	13%
	High school or less	26%	26%	28%
	Some college	20%	20%	23%
	College graduate	29%	30%	25%
	Post-graduate	10%	10%	10%
Race/Ethnicity	Non-Hispanic white	60%	66%	66%
	Non-Hispanic black	15%	12%	12%
	Hispanic	17%	15%	15%
	Other	8%	7%	7%
Marital Status	Married/partnered	49%	60%	52%
	Widowed/divorced/separated	27%	18%	20%
	Never married	24%	22%	28%
Employed	Yes	35%	33%	39%
	No	65%	67%	61%
Income	1st quartile	22%	15%	19%
	2nd quartile	20%	17%	20%
	3rd quartile	21%	22%	20%
	4th quartile	21%	28%	19%
	missing	17%	19%	22%

5.4 Results

Table 5.2 lists the estimates, standard errors (SE), root mean square error treating the correctly weighted NHIS as the true value (RMSE), percent-deviate from the NHIS estimate: %deviate = $100\left(\hat{T} - \hat{T}_y^{\text{NHIS}}\right)\Big/\hat{T}_y^{\text{NHIS}},$ and the standard deviation and minimum and maximum of the weights associated with a given estimator. We treat NHIS estimate as the unbiased estimate because it is calculated with probability-based sampling weights provided by NHIS. Without any weighting adjustment, HTSRS shows a positive bias of 5.9%. The GREG2 estimator reduces this bias from 5.9% to 2.0%, the GREG1 estimator reduces bias to 1.8%, while LASSO estimator reduces the bias to 0.9%. By definition, use of the model-assisted estimator using linear predictors will yield the same estimator as the GREG model; however the variability is substantially reduced. In this analysis, if NHIS were a non-probability sample, without weighting adjustment, we would have over-counted the number of adults with cancer by 1.18 million. With traditional calibration, the error is reduced to an over-count of 365 thousand (without variable selection) or 392 thousand (with variable selection). LASSO calibration further reduces the over-count to 175 thousand.

Table 5.2
Results for estimating total number of individuals with cancer. % deviate is the difference to NHIS estimate divided by the NHIS estimate

Estimator	$\hat{T}$	SE	RMSE	% deviate from NHIS	SD (min, max) of weights
NHIS	19,889,327	492,263	492,263	0.00%	5,913 (168; 93,244)
HTSRS	21,070,498	362,883	1,235,657	5.94%	0 (6,866; 6,866)
GREG1	20,254,449	375,064	523,438	1.84%	2,474 (-2,409; 16,679)
GREG1 MC	20,254,449	349,100	505,158	1.84%	269 (6,181; 7,326)
GREG2	20,281,603	367,900	537,802	1.97%	2,039 (-626; 13,947)
GREG2 MC	20,281,603	349,552	525,421	1.97%	260 (6,215; 7,291)
LASSO	20,064,671	347,586	389,309	0.88%	323 (5,786; 7,168)

As expected, the standard error of the NHIS estimate is the largest, as it properly incorporates complex survey design. If the calibration working model correctly captures the relationship between the outcome variable and the calibration variables, we anticipate that the calibration estimator standard errors to be smaller than HTSRS estimator's. This is not the case for either of the GREG estimator, where the standard error is larger than HTSRS's, although the RMSE is smaller due to the reduction in bias. In addition, the standard GREG estimator has a standard error about 2.0% greater than the backward selection GREG estimator, a feature offset by its estimated 6.6% reduction in bias (although this is insufficient to reduce RMSE); use of the model-assisted GREG estimator does reduce the standard error, and the root mean square error, by 5-7% and 2-3% respectively, over the standard GREG estimates. For LASSO calibration, we do observe a smaller standard error than HTSRS's, even with the bootstrap variance estimate that tends to overestimate. Without using the correct design weights, LASSO calibration produced the most accurate estimate of a population total while providing the smallest standard error among the estimators in this application. This is in spite of the fact that the standard deviation of the LASSO calibration weights were only about one-seventh as variable as the GREG weights, reflected in the smaller standard error of the estimator itself and greatly reduced RMSE.

6 Conclusion

In this manuscript, we developed the LASSO calibration estimator of population totals, $\hat{T}_y^{\mathrm{LASSO}}$, given population auxiliary data. We also derived closed-form variance estimates for $\hat{T}_y^{\mathrm{LASSO}}$. Simulation results show that the point estimates are approximately unbiased under simple-random sampling and informative sampling. For sample selections that are related to analysis variables, LASSO was able to significantly reduce sample bias even without the correct design weights. LASSO tends to outperform stepwise-selected working models when covariates are highly collinear. For analysis with many categorical variables, where there are natural correlations between the categories, LASSO calibration estimator can perform better than traditional calibration estimators, even if the effect sizes are small. The improvement is modest in the continuous variable setting, but substantial when the outcome of interest is binary, as shown in simulations and in the NHIS data example. We have demonstrated theoretically and through simulations that LASSO calibration holds great promise in making unbiased inference of population totals from non-probability samples. Although asymptotic closed-form variance estimates did not produce very accurate nominal coverage, the naive bootstrap is a viable alternative approach. In an application to estimate population total of individuals diagnosed with cancer, without correct design weights, the LASSO calibration estimator was able to produce an estimate that is the closest to the estimate based on correct survey weights. LASSO calibration estimator also has the smallest standard error of all the estimators considered, although the bootstrap variance estimate that was used did not fully account for the clustering in the NHIS, which generally increases standard errors. The application shows that LASSO calibration can generate inference to the population for a specific outcome variable, and the inference is both more accurate and precise than traditional calibration estimators.

The question arises when use of LASSO model-assisted calibration should be used instead of traditional calibration methods such as GREG. Both theoretical and empirical results in this paper suggest that there is little to be lost in terms of statistical efficiency to use LASSO model-assisted calibration, it does require additional effort on the part of the analyst to implement. While we cannot give specific cutoffs, our analysis suggests that this effort will be worthwhile when a) there are large numbers of potential calibration variables, b) many of these calibration variables are likely to be highly correlated, and c) the outcome is binary rather than continuous. We believe that conditions a) and b), at least, are increasing likely to be encountered in non-probability settings, where administrative datasets might provide these types of calibration variables and subsets of data obtained through various means will contain the core variable of interest.

While LASSO provides particularly convenient and rapid implementation, there are, of course, other modern regression methods that could be considered in addition to LASSO to develop penalized regression models for high-dimensional model-assisted regression, including approaches such as ridge regression, principle components, or Bayesian additive regression trees (Chipman, George and McCulloch, 2010). These approaches provide opportunity for further research in this area.

Finally, we note that this work is only a part of a larger and rapidly expanding literature on inference from non-probability surveys. In addition to the work of McConville et al. (2017), the “Mr. P” (multi-level regression and poststratification or MRP) approach of Wang et al. (2015) also uses high dimensional covariates to adjust non-probabilities samples, by use of a hierarchical model rather than penalized regression. Quasi-randomization (Elliott, 2009; Elliott, Resler, Flannagan and Rupp, 2010; Elliott and Valliant, 2017) and sample matching (Rivers, 2007; Vavreck and Rivers, 2008) also provide alternatives that use data from either known population quantities or probability sampling estimates to deal with selection bias issues in non-probability samples. Each have their strengths and weaknesses relative to each other and to model-assisted LASSO. The MRP approach makes distributional assumptions that might improve efficiency, but might reduce robustness, and is non-trivial to implement in its fully Bayesian form. Quasi-randomization forfeits the link to a particular outcome variable, making the weights it develops general purpose but likely less effective, while sample matching requires intervention at the design stage to sample elements from the non-probability frame that match elements from the population, ala quota sampling. The decision to use model-assisted LASSO calibration should be made in the context of these tradeoffs.

Acknowledgements

The authors thank Professor Fred M. Feinberg and Associate Research Scientist Sunghee Lee for their helpful review and suggestions, as well as the Editor, Associate Editor, and three referees whose suggestions greatly contributed to improving this paper.

Appendix

Determining estimates for adaptive LASSO

In practice, we do not observe the theoretical rate of growth of λ_n, which optimize model fit measures such as AIC or BIC, unless we have obtained many samples of the same population with various sample sizes. Given a sample, the choices of λ_n and γ depend on the modeler. In R *glmnet* implementation (Friedman et al., 2010), a range of λ_n is determined by the following scheme:

1. Set $\gamma = 0$.
2. Determine $\lambda_n^{\max}$ by finding the smallest λ_n that sets all coefficients to 0.
3. If sample size n is larger than the number of parameters in the regression model, set $\lambda_n^{\min} = 0.0001\lambda_n^{\max}$. If sample size n is smaller than the number of parameters, set $\lambda_n^{\min} = 0.01\lambda_n^{\max}$ (to set parameters to 0 sooner).
4. Generate a grid of λ_n, typically 100 equally spaced points between $\lambda_n^{\min}$ and $\lambda_n^{\max}$.

The initial range of values of λ_n is determined independently of γ. Choices of γ are less data-driven. Some modelers choose one of $\gamma = 0.1, 0.5, 1, 2$. Here we determine (λ_n, γ) through cross-validation as follows:

1. Obtain $\alpha_j = 1/|\hat{\beta}_j^{\text{MLE}}|$.
2. Determine 100 equally spaced values of λ_n based on R *glmnet* 's implementation.
3. For each pair (λ_n, γ), λ_n from Step 2, and $\gamma = 0.1, 0.5, 1, 2$, split data into 5 folds. Use 4 folds to obtain $\hat{\boldsymbol{\beta}}$.

4. Apply $\hat{\boldsymbol{\beta}}$ to the last fold not used to estimate $\hat{\boldsymbol{\beta}}$ and calculate a metric. For continuous $\mathbf{y}$, we calculate the mean-absolute-error (MAE), $\sum_{i \in s_{A(k)}} | \hat{\mu}_i - y_i |$. For binary $\mathbf{y}$, we calculate the area under curve (AUC) (calculated through R *glmnet* :: *auc* function).
5. Average the 5 metrics for each pair of (λ_n, γ), and choose the pair with the best average metric: minimum MAE for continuous $\mathbf{y}$, maximum AUC for binary $\mathbf{y}$.

The adaptive LASSO coefficient estimates are then obtained by solving equations (3.1) or (3.2) in Section 3.1 given the selected (λ_n, γ). The R code used to perform cross-validation is provided in the online supplemental material.

Asymptotic unbiasedness and variance of model-assisted LASSO calibration estimator of a population total

Lemma 1: *Assume the superpopulation model:*

$$E_\xi (y_k \,|\, \mathbf{x}_k) = \mu(\mathbf{x}_k, \boldsymbol{\beta}), V_\xi (y_k \,|\, \mathbf{x}_k) = v_k^2 \sigma^2.$$

Let $\mathbf{B}$ *be the finite-population quasilikelihood estimate of* $\boldsymbol{\beta}, \mathbf{B} \to \boldsymbol{\beta}$. *Under conditions (1)-(5) in Section 3.2, the model-assisted asymptotic estimator of population total is:*

$$\hat{T}_y^{MC} = \sum_{i \in s_A} d_i^A (y_i - \mu_i B^{MC}) + \sum_{i=1}^{N} \mu_i B^{MC} + o_p\left(\frac{N}{\sqrt{n}}\right) \qquad \text{(A.1)}$$

where

$$\mu_i = \mu(\mathbf{x}_i, \mathbf{B})$$

$$B^{MC} = \frac{\sum_{i=1}^{N} (\mu_i - \bar{\mu})(y_i - \bar{y})}{\sum_{i=1}^{N} (\mu_i - \bar{\mu})^2}.$$

Proof. The proof is adapted and expanded from the proof of Theorem 1 in Wu and Sitter (2001), with slight modifications in notations to be consistent with this paper. We begin by deriving the asymptotic model-assisted estimator for a population mean, $\hat{\bar{y}}^{\text{MC}} = N^{-1}\hat{T}_y^{\text{MC}}$ *(see equation (2.7)). By conditions (2) and (3), the second order Taylor series expansion of* $\mu(\mathbf{x}_i, \hat{\mathbf{B}})$ *around* $\mathbf{B}$ *is:*

$$\mu(\mathbf{x}_i, \hat{\mathbf{B}}) = \mu(\mathbf{x}_i, \mathbf{B}) + \left\{ \frac{\partial \mu(\mathbf{x}_i, \mathbf{t})}{\partial \mathbf{t}} \bigg|_{\mathbf{t}=\mathbf{B}} \right\}^T (\hat{\mathbf{B}} - \mathbf{B}) + (\hat{\mathbf{B}} - \mathbf{B})^T \left\{ \frac{\partial^2 \mu(\mathbf{x}_i, \mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^T} \big|_{\mathbf{t}=\mathbf{B}^*} \right\} (\hat{\mathbf{B}} - \mathbf{B}) \qquad \text{(A.2)}$$

for $\mathbf{B}^* \in (\hat{\mathbf{B}}, \mathbf{B})$ *or* $(\mathbf{B}, \hat{\mathbf{B}})$. *Let*

$$\mathbf{h}(\mathbf{x}_i, \mathbf{B}) = \frac{\partial \mu(\mathbf{x}_i, \mathbf{t})}{\partial \mathbf{t}} \big|_{\mathbf{t}=\mathbf{B}}$$

$$\mathbf{k}(\mathbf{x}_i, \mathbf{B}^*) = \frac{\partial^2 \mu(\mathbf{x}_i, \mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^T} \bigg|_{\mathbf{t}=\mathbf{B}^*}$$

Note that $\mathbf{h}$ *is a vector of length* m *and* $\mathbf{k}$ *is a matrix of size* $m \times m,$ *where* m *is the number of parameters in* $\boldsymbol{\beta}$. *By conditions (2) and (3),*

$$\max_i \left|\mathbf{h}(\mathbf{x}_i,\, \mathbf{B})\right| \;\le\; h(\mathbf{x}_i,\, \mathbf{B}) \tag{A.3}$$

$$\max_{k,\, j} \left|\mathbf{k}(\mathbf{x}_i,\, \mathbf{B}^*)\right| \;\le\; k(\mathbf{x}_i,\, \mathbf{B}^*). \tag{A.4}$$

Conditions (1) and (3) imply that

$$\mu\left(\mathbf{x}_i,\, \hat{\mathbf{B}}\right) = \mu(\mathbf{x}_i,\, \mathbf{B}) + O_p\left(1\big/\sqrt{n}\right) \tag{A.5}$$

$$\equiv \mu_i + O_p\left(1\big/\sqrt{n}\right). \tag{A.6}$$

By equation (2.2) in Section 2.1 and the boundedness conditions of (2) and (3) in Section 3.2.2 imply

$$\begin{aligned} N^{-1}\sum_{i\in s_A} d_i^A \mu\left(\mathbf{x}_i,\, \hat{\mathbf{B}}\right) &= N^{-1}\sum_{i\in s_A} d_i^A \mu(\mathbf{x}_i,\, \mathbf{B}) + N^{-1}\left(\sum_{i\in s_A} d_i^A \mathbf{h}(\mathbf{x}_i,\, \mathbf{B})\right)^T \left(\hat{\mathbf{B}} - \mathbf{B}\right) \\ &\quad + \left(\hat{\mathbf{B}} - \mathbf{B}\right)^T N^{-1}\left(\sum_{i\in s_A} d_i^A \mathbf{k}(\mathbf{x}_i,\, \mathbf{B}^*)\right)\left(\hat{\mathbf{B}} - \mathbf{B}\right) \\ &= N^{-1}\sum_{i\in s_A} d_i^A \mu(\mathbf{x}_i,\, \mathbf{B}) + N^{-1}\left(\sum_{i\in s_A} d_i^A \mathbf{h}(\mathbf{x}_i,\, \mathbf{B})\right)^T \left(\hat{\mathbf{B}} - \mathbf{B}\right) + O_p\left(\frac{1}{n}\right). \end{aligned} \tag{A.7}$$

By conditions (1), (4), and equation (A.7):

$$\begin{aligned} N^{-1}\sum_{k=1}^{N} \mu\left(\mathbf{x}_k,\, \hat{\mathbf{B}}\right) &- N^{-1}\sum_{i\in s_A} d_i^A \mu\left(\mathbf{x}_i,\, \hat{\mathbf{B}}\right) \\ &= N^{-1}\sum_{k=1}^{N} \mu(\mathbf{x}_i,\, \mathbf{B}) - N^{-1}\sum_{i\in s_A} d_i^A \mu(\mathbf{x}_i,\, \mathbf{B}) + O_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{A.8}$$

Using conditions (1) and (3),

$$\begin{aligned} \bar{\hat{\mu}} &= \sum_{i\in s_A} d_i^A \mu\left(\mathbf{x}_i,\, \hat{\mathbf{B}}\right) \Big/ \sum_{i\in s_A} d_i^A \\ &= \left(\sum_{i\in s_A} d_i^A\right)^{-1} \sum_{i\in s_A} d_i^A \left(\mu(\mathbf{x}_i,\, \mathbf{B}) + \mathbf{h}^T(\mathbf{x}_i,\, \mathbf{B})\left(\hat{\mathbf{B}} - \mathbf{B}\right) + \left(\hat{\mathbf{B}} - \mathbf{B}\right)^T \mathbf{k}(\mathbf{x}_i,\, \mathbf{B}^*)\left(\hat{\mathbf{B}} - \mathbf{B}\right)\right) \\ &= \left(\sum_{i\in s_A} d_i^A\right)^{-1} \sum_{i\in s_A} d_i^A \left(\mu(\mathbf{x}_i,\, \mathbf{B}) + \mathbf{h}^T(\mathbf{x}_i,\, \mathbf{B})\left(\hat{\mathbf{B}} - \mathbf{B}\right)\right) + O_p(1/n) \\ &= \bar{\mu} + \left(\sum_{i\in s_A} d_i^A\right)^{-1} \sum_{i\in s_A} d_i^A \mathbf{h}^T(\mathbf{x}_i,\, \mathbf{B})\left(\hat{\mathbf{B}} - \mathbf{B}\right) + O_p(1/n) \\ &\quad (\textit{by condition } (1) \textit{ and } (18)) \\ &= \bar{\mu} + O_p\left(1\big/\sqrt{n}\right) + O_p(1/n) \\ &= \bar{\mu} + O_p\left(1\big/\sqrt{n}\right) \end{aligned} \tag{A.9}$$

for $\bar{\mu} = \sum_{i\in s_A} d_i^A \mu_i \Big/ \sum_{i\in s_A} d_i^A.$

Then from (A.2) and (A.9) and using conditions (1)-(3), we have

$$\begin{aligned} N^{-1}\sum_{i\in s_A} d_i^A\left(\hat{\mu}_i - \hat{\bar{\mu}}\right) &= N^{-1}\sum_{i\in s_A} d_i^A\Big(\mu\left(\mathbf{x}_i, \mathbf{B}\right) + \mathbf{h}^T\left(\mathbf{x}_i, \mathbf{B}\right)\left(\hat{\mathbf{B}} - \mathbf{B}\right) + \left(\hat{\mathbf{B}} - \mathbf{B}\right)^T \mathbf{k}\left(\mathbf{x}_i, \mathbf{B}^*\right)\left(\hat{\mathbf{B}} - \mathbf{B}\right) - \bar{\mu}\Big) \\ &= N^{-1}\sum_{i\in s_A} d_i^A\left(\mu_i - \bar{\mu}\right) + N^{-1}\sum_{i\in s_A} d_i^A \mathbf{h}^T\left(\mathbf{x}_i, \mathbf{B}\right)\left(\hat{\mathbf{B}} - \mathbf{B}\right) \\ &\quad + N^{-1}\sum_{i\in s_A} d_i^A\left(\hat{\mathbf{B}} - \mathbf{B}\right)^T \mathbf{k}\left(\mathbf{x}_i, \mathbf{B}^*\right)\left(\hat{\mathbf{B}} - \mathbf{B}\right) - O_p\left(1/\sqrt{n}\right) \\ &= N^{-1}\sum_{i\in s_A} d_i^A\left(\mu_i - \bar{\mu}\right) + O_p\left(1/\sqrt{n}\right) + O_p\left(1/n\right) - O_p\left(1/\sqrt{n}\right) \\ &= N^{-1}\sum_{i\in s_A} d_i^A\left(\mu_i - \bar{\mu}\right) + O_p\left(1/\sqrt{n}\right). \end{aligned} \tag{A.10}$$

Similarly,

$$N^{-1}\sum_{i\in s_A} d_i^A\left(\hat{\mu}_i - \hat{\bar{\mu}}\right)^2 = N^{-1}\sum_{i\in s_A} d_i^A\left(\mu_i - \bar{\mu}\right)^2 + O_p\left(1/n\right). \tag{A.11}$$

From (A.10) and (A.11) we have:

$$\begin{aligned} \hat{B}^{MC} &= \frac{\sum_{i\in s_A} d_i^A\left(\hat{\mu}_i - \hat{\bar{\mu}}\right)\left(y_i - \bar{y}\right)}{\sum_{i\in s_A} d_i^A\left(\hat{\mu}_i - \hat{\bar{\mu}}\right)^2} = \frac{N^{-1}\sum_{i\in s_A} d_i^A\left(\hat{\mu}_i - \hat{\bar{\mu}}\right)\left(y_i - \bar{y}\right)}{N^{-1}\sum_{i\in s_A} d_i^A\left(\hat{\mu}_i - \hat{\bar{\mu}}\right)^2} \\ &= \frac{\sum_{i\in s_A} d_i^A\left(\mu_i - \bar{\mu}\right)\left(y_i - \bar{y}\right) + O_p\left(1/\sqrt{n}\right)}{\sum_{i\in s_A} d_i^A\left(\mu_i - \bar{\mu}\right)^2 + O_p\left(1/n\right)} \\ &\to B^{MC} \qquad as\ n \to \infty. \end{aligned} \tag{A.12}$$

Thus $\hat{B}^{MC} = B^{MC} + o_p\left(1\right)$, *and we have:*

$$\begin{aligned} \hat{\bar{y}}^{MC} &= N^{-1}\hat{T}_y^{MC} \\ &= N^{-1}\mathbf{d}^A\mathbf{y} + \left(N^{-1}\sum_{k=1}^{N}\mu\left(\mathbf{x}_k, \hat{\mathbf{B}}\right) + \sum_{i\in s_A} N^{-1} d_i^A \mu\left(\mathbf{x}_i, \hat{\mathbf{B}}\right)\right)\hat{B}^{MC} \\ &= N^{-1}\mathbf{d}^A\mathbf{y} + \left(N^{-1}\sum_{k=1}^{N}\mu\left(\mathbf{x}_k, \mathbf{B}\right) - N^{-1}\sum_{i\in s_A} d_i^A \mu\left(\mathbf{x}_i, \mathbf{B}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right)\left(B^{MC} + o_p\left(1\right)\right) \\ &= N^{-1}\mathbf{d}^A\mathbf{y} + \left(N^{-1}\sum_{k=1}^{N}\mu\left(\mathbf{x}_k, \mathbf{B}\right) - N^{-1}\sum_{i\in s_A} d_i^A \mu\left(\mathbf{x}_i, \mathbf{B}\right)\right)B^{MC} + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Since $N = O_p\left(N\right)$, *we have* $N \cdot o_P\left(1/\sqrt{n}\right) = O_p\left(N\right)o_p\left(1/\sqrt{n}\right) = o_p\left(N/\sqrt{n}\right)$. *Thus,*

$$\begin{aligned} \hat{T}_y^{MC} &= N\hat{\bar{y}}^{MC} = N\left(N^{-1}\mathbf{d}^A\mathbf{y} + \left(N^{-1}\sum_{k=1}^{N}\mu\left(\mathbf{x}_k, \mathbf{B}\right) - N^{-1}\sum_{i\in s_A}\mu\left(\mathbf{x}_i, \mathbf{B}\right)\right)B^{MC} + o_p\left(\frac{1}{\sqrt{n}}\right)\right) \\ &= \mathbf{d}^A\mathbf{y} + \left(\sum_{k=1}^{N}\mu\left(\mathbf{x}_k, \mathbf{B}\right) - \sum_{i\in s_A}\mu\left(\mathbf{x}_i, \mathbf{B}\right)\right)B^{MC} + o_p\left(\frac{N}{\sqrt{n}}\right) \\ &= \sum_{iins_A} d_i^A\left(y_i - \mu_i B^{MC}\right) + \sum_{i=1}^{N}\mu_i B^{MC} + o_p\left(\frac{N}{\sqrt{n}}\right). \end{aligned} \tag{A.13}$$

Theorem 2: *Suppose the parameters in a full regression model have both zero and non-zero components, without loss of generality, let the first* p *be non-zero and the last* q *be zero:*

$$\boldsymbol{\beta}^F = \begin{pmatrix} \boldsymbol{\beta}^{(1)}_{(p \times 1)} \\ \boldsymbol{\beta}^{(2)}_{(q \times 1)} \end{pmatrix}, \ \boldsymbol{\beta}^{(1)} = \boldsymbol{\beta} \quad and \quad \boldsymbol{\beta}^{(2)} = \mathbf{0}_{(q \times 1)}.$$

Under conditions (1)-(5), the asymptotic LASSO calibration estimator of total is:

$$\hat{T}_y^{LASSO} = \sum_{i \in s_A} d_i^A \left(y_i - \mu_i B^{MC}\right) + \sum_{i=1}^{N} \mu_i B^{MC} + o_p\left(\frac{N}{\sqrt{n}}\right). \tag{A.14}$$

Proof. Under condition (5), the adaptive LASSO regression satisfies the oracle property through Theorems 1 and 4 in Zou (2006):

$$\begin{aligned} Pr\left(\mathbf{B}^{(2)} = \mathbf{0}\right) &\rightarrow 1 \\ \sqrt{n}\left(\hat{\mathbf{B}}^{(1)} - \mathbf{B}\right) &\rightarrow N\left(\mathbf{0}, \mathbf{C}\right) \\ \mathbf{B} &\rightarrow \boldsymbol{\beta} \end{aligned}$$

where $\mathbf{C} = \Sigma\left(\mathbf{B}\right)$ *is the covariance matrix of* $\mathbf{B}^{(1)}$ *under the linear model, and* $\mathbf{C} = I^{-1}\left(\mathbf{B}\right)$ *is the inverse of Fisher information matrix of* $\mathbf{B}^{(1)}$ *under generalized linear model. By Slutsky's theorem, the oracle property implies* $\hat{\mathbf{B}}^{(1)} = \mathbf{B} + O_p\left(1/\sqrt{n}\right)$. *By condition (1) and Lemma 1:*

$$\begin{aligned} \hat{T}_y^{LASSO} &\approx \hat{T}_y^{MC} \\ &= \sum_{iins_A} d_i^A \left(y_i - \mu_i B^{MC}\right) + \sum_{i=1}^{N} \mu_i B^{MC} + o_p\left(\frac{N}{\sqrt{n}}\right). \end{aligned}$$

Theorem 3: $\hat{T}_y^{LASSO}$ *is model-unbiased.*

Proof. Under the assumption of our theoretical framework, the superpopulation parameters are a subset of the full LASSO regression parameters, we can prove the model-unbiasedness of $\hat{T}_y^{LASSO}$ *by taking expectations with respect to model* ξ. *First note that:*

$$E_\xi\left[B^{MC}\right] = E_\xi\left[\frac{\sum_{i=1}^{N}\left(\mu_i - \bar{\mu}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{N}\left(\mu_i - \bar{\mu}\right)^2}\right] = \frac{\sum_{i=1}^{N}\left(\mu_i - \bar{\mu}\right)\left(\mu_i - \bar{\mu}\right)}{\sum_{i=1}^{N}\left(\mu_i - \bar{\mu}\right)^2} = 1.$$

Thus

$$\begin{aligned} E_\xi\left[\hat{T}_y^{LASSO} - T\right] &\approx E_\xi\left[\sum_{i \in s_A} d_i^A \left(y_i - \mu_i B^{MC}\right) + \sum_{i=1}^{N} \mu_i B^{MC} - \sum_{i=1}^{N} y_i\right] \\ &= \sum_{i \in s_A} d_i^A \left(\mu_i - \mu_i\right) + \sum_{i=1}^{N} \mu_i - \sum_{i=1}^{N} \mu_i \qquad \left(since\ E_\xi\left[B^{MC}\right] = 1\right) \\ &= 0. \end{aligned}$$

Thus, as long as LASSO regression parameters include the superpopulation parameters, $\hat{T}_y^{\text{LASSO}}$ is model-unbiased regardless of design weights. This property is essential in non-probability samples, where there are no initial design weights to guarantee unbiasedness.

References

Baker, R., Brick, J.M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K. and Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1, 90-143.

Chipman, H.A., George, E.I. and McCulloch, R.E. (2010). BART: Bayesian additive regression trees. *Annals of Applied Statistics*, 4, 266-298.

Cardot, H., Goga, C. and Shehzad, M.-A. (2017). Calibration and partial calibration on principal components when the number of auxiliary variables is large. *Statistica Sinica*, 27, 243-260.

Centers for Disease Control and Prevention (2005). *2004 National Health Interview Survey (NHIS) Public Use Data Release: NHIS Survey Description*. National Center for Health Statistics: Hyattsville, Maryland. www.cdc.gov/nchs/data/nhis/srvydesc.pdf.

Czanner, G., Sarma, S.V., Eden, U.T. and Brown, E.N. (2008). A signal-to-noise ratio estimator for generalized linear model systems. *Proceedings of the World Congress on Engineering*, vol. 2.

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

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Marquez, J.R.G., Gruber, B., Lafourcade, B., Leitao, P.J. and Mnkemller, T. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecology*, 36, 27-46.

Elliott, M.R. (2009). Combining data from probability and nonprobability samples using pseudo-weights. *Survey Practice*, 2(6).

Elliott, M.R, Resler, A., Flannagan, C. and Rupp, J. (2010). Combining data from probability and non-probability samples using pseudo-weights. *Accident Analysis and Prevention*, 42, 530-539.

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

Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348-1360.

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

Frankel, M.R., and Frankel, L.R. (1987). Fifty years of survey sampling in the United States. *Public Opinion Quarterly*, S127-S138.

Fuller, W.A. (2009). *Sampling Statistics*. New York: John Wiley & Sons, Inc.

Goga, C., Muhammad-Shehzad, A. and Vanheuverzwyn, A. (2011). Principal component regression with survey data: Application on the French media audience. *Proceedings of the 58th World Statistics Congress of the International Statistical Institute*, 3847-3852.

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70, 646-675.

Kamarianakis, Y., Shen, W. and Wynter, L. (2012). Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. *Applied Stochastic Models in Business and Industry*, 28, 297-315.

Kohannim, O., Hibar, D.P., Stein, J.L., Jahanshad, N., Hua, X., Rajagopalan, P., Toga, A., Jack Jr, C.R., Weiner, M.W., de Zubicaray, G.I. and McMahon, K.L. (2012). Discovery and replication of gene influences on brain structure using LASSO regression. *Frontiers in Neuroscience*, 6, 115.

Kohut, A., Keeter, S., Doherty, C., Dimock, M. and Christian, L. (2012). Assessing the representativeness of public opinion surveys. *Pew Research Center for The People & The Press*. http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/.

Mosteller, F. (1949). *The Pre-Election Polls of 1948: The Report to the Committee on Analysis of Pre-Election Polls and Forecasts*, vol. 60, Social Science Research Council.

McConville, K. (2011). *Improved Estimation for Complex Surveys Using Modern Regression Techniques*. Unpublished PhD Thesis, Colorado State University.

McConville, K., Breidt, F.J., Lee, T.M. and Moisen, G.G. (2017). Model-assisted survey regression estimation with the LASSO. *Journal of Survey Statistics and Methodology*, 5, 131-158.

Park, M., and Yang, M. (2008). Ridge regression estimation for survey samples. *Communication in Statistics - Theory and Methods*, 37, 532-543.

Rivers, D. (2007). Sampling for web surveys. *Proceedings of the Joint Statistical Meetings*, American Statistical Association.

Särndal, C.-E., Swensson, B. and Wretman, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76, 527-537.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Skinner, C., and Silva, P. (1997). Variable selection for regression estimation in the presence of nonresponse. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 76-81.

Stephan, F.F. (1948). History of the uses of modern sampling procedures. *Journal of the American Statistical Association*, 43, 12-39.

Terhanian, G., and Bremer, J. (2012). A smarter way to select respondents for surveys? *International Journal of Market Research*, 54, 751-780.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society*, 58, 267-288.

Tourangeau, R., Conrad, F.G. and Couper, M.P. (2013). *The Science of Web Surveys.* Oxford University Press, Oxford, UK.

Vavreck, L., and Rivers, D. (2008). The 2006 Cooperative Congressional Election Study. *Journal of Elections, Public Opinion, and Parties*, 355-366.

Wang, W., Rothschild, D., Goel, S. and Gelman, A. (2015). Forecasting elections with non-representative Polls. *International Journal of Forecasting*, 31, 980-991.

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

Wu, T.T., Chen, Y.F., Hastie, T., Sobel, E. and Lange, K. (2009). Genome-wide association analysis by LASSO penalized logistic regression. *Bioinformatics*, 25, 714-721.

Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, 101, 1418-1429.