Literature Review: Survey Weight Diagnostic Tests

Corbin Lubianski

January 5, 2024

1 Survey Weights

Survey data typically comprises of complex sample designs involving unequal sampling probabilities of population units, rather than simple random sampling (SRS) or other equal probability selection method (EPSEM) designs. Departing from equal probability selection methods require adjustments to account for complex sampling designs, unit nonresponse, frame coverage errors, and other sources of sampling bias. Survey statistics employ survey weights to avoid the bias of population means or proportions. However, the current literature is not decisive on extrapolating survey weights beyond calculating population statistics to more inferential statistics like regression coefficients.

Furthermore, the usage of survey weights is not deterministic of analytical results, but more merely tradition and field practices. From the experience of Bollen *et al.* (2016), they find biostatistics, public health, and federal statistics scientists generally use weights. Alternatively, they note fields that do not utilize weights typically comprise of social sciences like econometrics. [2] While it is easy for statisticians to claim social scientists naively ignore survey weights, accessible and easy-to-use survey weight diagnostic tests are nonexist to analysts.

1.1 What are Survey Weights?

Let $\pi_{Si} = \mathbb{P}(i \in S)$ be the selection probability of surveying observation i from sample S. The classic Horvitz-Thompson (HT) estimator of the population total Y is defined as

$$\hat{Y} = \sum_{i \in S} w_{Si} y_i \text{ where weight } w_{Si} = \pi_{Si}^{-1}.$$

In case of non-response, the HT estimator can be further generalized by replacing w_{Si} with $w_i = \pi_i^{-1}$ with $\pi_i = \mathbb{P}(i \in S, i \in F, i \in R)$ with F being the units in the target population within the sampling frame and R being the units in S that respond to the survey. Intuitively, the sampling weight of sampled unit i can be interpreted as the number of population units represented by unit i. [8]

Since survey weights are the design weights that projects sample-level statistics to the population, we can estimate the population size as

$$\hat{N} = \sum_{i \in S} w_i.$$

The population total of some characteristic y is estimated by

$$\hat{t} = \sum_{i \in S} w_i y_i,$$

and the population mean of y is estimated by

$$\bar{y} = \frac{\sum_{i \in S} w_i y_i}{\sum_{i \in S} w_i} = \frac{\hat{t}}{\hat{N}}.$$

1.2 Why use Survey Weights?

With sample data with unequal sampling probabilities from the population, using survey weights corrects the sample to become representative of the population. Survey weights reduce the bias of population-level estimators like population mean and proportions. However, a notable disadvantage of survey weights is that they can substantially inflate the variance of the model parameter estimates. As such, it may result in higher efficiency of estimators and statistical power when not opting for survey weights. Commonly, the Bias-Variance trade-off appears when considering survey weights.

1.2.1 Survey Weights and Regression

Linear regressions are a common model for scientists to find relationships between dependent and independent variables. As mentioned prior, using survey weights for linear regressions, even generalized linear models, is a highly debated topic within survey statistics. As discussed in Kish and Frankel (1974), survey statisticians acknowledge how various survey designs affect regression inference such as unequal weights affecting regression coefficients and survey design affecting coefficient standard errors. [7] Take the four main assumptions of a linear regression:

1. Linearity: $E(\varepsilon_i \mid X_i) = 0$, for all i;
2. Homoscedasticity: $\text{Var}(\varepsilon_i \mid X_i) = \sigma^2$, for all i;
3. Independence between observations: $\text{Cov}(\varepsilon_i, \varepsilon_j \mid \mathbf{X_i}, \mathbf{X_j}) = 0$, for all $i \neq j$;
4. Normality for ε_i.

Assumptions (1), (2), and (4) form the basic structure of the model while (3) relates to the selection procedure of the sample. Complex selection designs like stratified and cluster selection violates the "independence between observations" assumption. Specifically, complex selection designs introduce positive correlations between errors ε_i of the model which may translate to biased estimators and underestimated standard errors. [7]

1.2.2 Model-Based and Design-Based Inference

Before continuing, it's important to address the different population perspectives taken in regression inference as it is a common source of confusion regarding weights in inferential analysis. Concisely, the model-based perspective is regarded as the infinite population approach while the design-based perspective is regarded as the finite population approach. Although large-sample inferences of the two perspectives are often similar, the assumptions become notable in smaller sample sizes that typically coincide with complex survey data.

1. **Model-based Inference**. Consider a stochastic model that describes the relation between y_i and x_i that holds for *every observation in the population.* One common model which you could claim that Y_i is generated from x_i, per the following assumptions, is

   $$Y_i \mid \mathbf{x_i} = \mathbf{x}_i^T \beta + \varepsilon_i,$$

   where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ and independent of $\mathbf{X}$. If the observations in the population actually follow the model, then the sample design does not have any effect as long as the inclusion probabilities depend only on y. Some researchers may relax some assumptions such as replacing $\varepsilon \perp\!\!\!\perp \mathbf{X}$ with that they are only uncorrelated such that $\text{Cov}(\varepsilon, \mathbf{X}) = 0$.

   The model-based approach is attractive for regressions as it easily relates other statistical areas and is flexible in attributing survey designs aspects like accounting for nonresponse. Additionally, the model-based estimates can be used with relatively small samples and nonprobability samples. Under the aforementioned assumptions, the ordinary least squares (OLS) estimator of β is unbiased and consistent. [8]

2. **Design-based Inference**. The design-based approach is interested in the finite population characteristics $\mathbf{B}$ in the model

$$\mathbf{Y}_N = \mathbf{X}_N^T\mathbf{B} + \mathbf{E}_N,$$

where $\mathbf{Y}_N$ is an $N \times 1$ vector, $\mathbf{X}_N$ is an $N \times k$ matrix of explanatory variables with the first column as regression coefficients, $\mathbf{B}$ as the $k \times 1$ vector of coefficients, and $\mathbf{E}_N$ is an $N \times 1$ vector of errors. The finite population OLS estimator is

$$\mathbf{B} = (\mathbf{X}_N^T\mathbf{X}_N^T)^{-1}\mathbf{X}_N^T\mathbf{Y}_N.$$

No assumptions are made about how well the model fits the population nor the distribution of the residual errors. Inferences are based on repeated sampling from the finite population. Design-based analysts decide to fit a particular model if they believe it is a plausible candidate in describing the population. Since all inference is based on the survey design, *survey weights are required* to estimate the parameters and the survey design to estimate the variance of the regression coefficients.

3. **Model-*assisted* Inference**. As a combination of design-based and model-based approaches, Särndal *et al.* (1992) proposed a model-assisted approach that is assumed to be generated from a finite population, yet all inference is based on the survey design. Thus, the survey weights are used to calculate the regression coefficients $\hat{\mathbf{B}}$ and standard errors are calculated by the survey design. This produces a consistent estimator

$$\hat{\mathbf{B}} = (\mathbf{X}_n^T\mathbf{X}_n^T)^{-1}\mathbf{X}_n^T\mathbf{Y}_n$$

if $\mathrm{Cov}(w_j, y_j \mid \mathbf{x}_j) = 0$ for all j. [14] Acknowledging this other approach is imperative to understanding most of the weighting tests as much of the surrounding literature utilize this perspective, either explicitly or inferred. Bollen *et al.* (2016) notes that this a source of misunderstanding sample weights as "much of the confusion in the literature stems from a lack of specificity about which perspective is assumed for a test."

1.3 Poststratification versus Sampling Probabilities

Gelman (2007) argues that contrary to what is assumed by many theoretical statistics, survey weights do not generally equal the inverse selection probabilities and rather are typically based on more complex combination of other selection probabilities and non-response. [5] He suggests poststatification as a strategy for correcting a multitude of differences between a sample and the population.

1.3.1 Poststratification

The purpose of poststratification is to correct for differences between groups that are not known prior to sampling. With variables $\mathbf{X}$ whose joint distribution within the population is known, we can label possible categories of $\mathbf{X}$ as poststratification cells j, when assuming $\mathbf{X}$ is discrete, with population sizes N_j and sample sizes n_j. In general, poststratification subjects a sampled population to stratifying which then models the poststrata J as a simple random sample. As such, the population mean of the poststrata population means $\bar{y}_{\mathcal{U},j}$ is

$$\bar{y}_{\mathcal{U}} = \frac{\sum_{j=1}^{J} N_j \bar{y}_{\mathcal{U},j}}{\sum_{j=1}^{J} N_j}.$$

Since survey weights are not attributes to the observational units but rather constructions based on the entire survey, the survey weights within each poststrata J have the same poststratification weights. For each factors associated with poststratification, the product of the dimensions results in the size of the poststratified categories. It is important to choose these wisely considering extraneous factor interactions can quickly jump in size where some poststrata might have sizes of 1 or 0.

1.3.2 Sampling Probabilities

As the traditional route of constructing survey weights, using the sampling probabilities to construct survey weights are used during design phase of the survey since they are known prior to sampling and are normally fixed. They compensate for unequal probability of selection and unequal response rates and control the proportional contribution of each observational unit to the overall population. Constructing survey weights using sampling probabilities typically uses a few variables, but also need to be chosen strategically to avoid any covariance between the weights and the variables of interest.

2 Motivation

The motivation for research in survey weight diagnostic tests stem from a recent review by Bollen *et al.* (2016) to which they classify the existing literature of survey weight diagnostic tests for determining their necessity into two groups. The authors find that nearly all weighting tests fall into two categories: difference-in-coefficients (DC) tests and weight-association tests (WA). Bollen *et al.* (2016) conceptually reviewed the assumptions and properties of the tests and noted the limitations of existing Monte Carlo simulation studies about the tests. [2]

Beyond reviewing the existing literature, Bollen *et al.* (2016) noted unaddressed questions that remain from survey weight diagnostic test literature. This includes:

- To what degree do the asymptotic properties of the WA and DC tests become reliable with determining the necessity of including survey weights?
- In what situations are DC or WA tests interchangeable and when is one favored to the other?
- How do researchers determine the necessity of survey weights when there are a multitude of variables? Can researchers selectively choose which variables to test the weights?
- Beyond ordinary least squares (OLS), how do the tests perform in other models, like generalized linear models (GLM)?
- How do the Type I and Type II rejection rates compare between WA and DC tests?
- While most simulations handle simple unequal selection probabilities, how do different complex sampling affect the diagnostic tests considering most surveys are complexly designed?

Beyond Bollen *et al.* (2016), survey weight diagnostic test literature revolve around proposing individual tests and limited simulation studies to suggest their usability in analysis.

3 Diagnostic Tests

Bollen *et al.* (2016) main contributions are their collection of survey weight diagnostic tests and classifying them into two categories: weight-association tests and difference-in-coefficients tests. The article concludes by establishing the asymptotic equivalence between weight association and difference-in-coefficients tests.[2] Before reviewing the diagnostic tests, Bollen *et al.* (2016) iterated the importance of understanding the different approaches of survey statistics: design-based and model-based. They combined the perspectives by suggesting to regard the finite population of the design-based perspective as a simple random simple from the superpopulation described in the model-based perspective. Refer to Section 1.2.2 for more details. Recall that the assumption needed to infer based on the weights and dependent variables is

$$\mathrm{Cov}(w_j, y_j \mid \mathbf{x}_j) = 0.$$

3.1 Difference-in-Coefficients Tests

The difference-in-coefficients tests (DC) compare the coefficients of the weighted and unweighted regressions to determine whether the coefficient differences are statistically significantly different from zero. Starting with

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \text{ assuming } E(\varepsilon \mid \mathbf{X}) = 0 \text{ and } \mathrm{Var}(\varepsilon \mid \mathbf{X}) = \sigma^2\mathbf{I}.$$

Hausman (1978) create the basis of the DC test as a test for general misspecifications. Hausman proposed two linear regressions which output two equally-sized estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ of the β estimators. In a correctly specified model, the asymptotic value of $(\hat{\beta}_1 - \hat{\beta}_2)$ should be zero. Otherwise, if misspecification exists, then $(\hat{\beta}_1 - \hat{\beta}_2)$ should be nonzero. Hausman's proposed test statistic T_H is

$$T_H = (\hat{\beta}_1 - \hat{\beta}_2)' \hat{V}_H^{-1} (\hat{\beta}_1 - \hat{\beta}_2)$$

where $\hat{V}_H = \hat{V}(\hat{\beta}_1) - \hat{V}(\hat{\beta}_2)$ as the estimator of the asymptotic covariance matrix. Lastly, $T_H \sim \chi^2_{dim(\hat{\beta})}$ with degrees of freedom equal to the dimension of $\hat{\beta}$. [6]

3.1.1 Hausman-Pfeffermann DC Test

Pfeffermann (1993) proposed using the Hausman test for misspecification as a test to compare the coefficients of weighted and unweighted regressions as $\hat{\beta}_1 = \hat{\beta}_w$ referring to the coefficients of the weighted regression and $\hat{\beta}_2 = \hat{\beta}_u$ as the coefficients of the unweighted regression. This also corresponds with the covariance matrix estimator $\hat{V} = \hat{V}(\hat{\beta}_w) - \hat{V}(\hat{\beta}_u)$. [9]

A notable issue with this test statistic is the event the covariance estimator is negative which could corresponds to a negative chi-squared test statistic. As probability theory defines variance of random variables to be nonnegative, Hausman (1993) proposed this covariance estimator since under the null hypothesis, $\mathrm{Cov}(\hat{\beta}_u, \hat{\beta}_w - \hat{\beta}_u) = 0$. Unfortunately, this estimator is not necessarily positive definite, especially for small and moderate sample sizes when $\hat{\beta}_w$ will inflate as noted within the literature. Our next test will address this issue.

3.1.2 Asparouhov-Muthen DC Test

Asparouhov & Muthen (2007) extended the Hausman-Pfeffermann test by proposing a different estimator for V that is always positive definite. [1] Specifically, they proposed

$$\hat{V}_{AM} = \hat{V}(\hat{\beta}_w) + \hat{V}(\hat{\beta}_u) - 2C$$

where C is an estimator of the covariance matrix of the two estimators as

$$C = \left(\frac{\partial^2 L_1(\hat{\beta}_{w_1})}{(\partial\beta)^2} \right)^{-1} M \left(\frac{\partial^2 L_1(\hat{\beta}_{w_1})}{(\partial\beta)^2} \right)^{-1'}$$

$$M = \sum_i w_{1,i} w_{2,i} \frac{\partial l_i(\hat{\beta}_{w_1})}{\partial\beta} \left(\frac{\partial l_i(\hat{\beta}_{w_2})}{\partial\beta} \right)' .$$

3.2 Weight-Association Tests

The basis for many of weight-association (WA) tests also stem from Hausman (1978) misspecification tests where Hausman wanted to assess the statistical significance of β_M in the equation

$$Y = X\beta + X_M\beta_M + \varepsilon$$

where X_M is the transformed version of X. The null hypothesis is $H_0 : \beta_M = 0$ such that the regression coefficients of the weighted explanatory variables is non-information of Y. An F-test of H_0 requires an additional assumption that ε is Normally distributed.

3.2.1 DuMouchel-Duncan WA Test

Although Hausman only specified this form as a misspecification test, DuMouchel & Duncan (1983) extended the test to determine the decision of weighting in regressions. With regards to weights, a WA test checks whether

$$H_0 : E(Y \mid X, W) = E(Y \mid X).$$

Within this context, consider the regression

$$Y = X\beta_u + X_w\beta_w + \varepsilon.$$

DuMouchel & Duncan (1983) recommend estimating the regression model with ordinary least squares (OLS) and then testing the null hypothesis of $H_0 : \beta_w = 0$ using an F-test to determine whether weights are needed in the analysis. [4]

3.2.2 Pfeffermann-Sverchkov WA Test (1999)

Pfeffermann and Sverchkov proposed multiple WA tests in a sequence of works. Pfeffermann & Sverchkov (1999) derived several tests where they investigate the relationships between the sample unweighted residuals and weights in a regression. They argue that if the sample distribution of the residuals is the same as the population distribution, then you can ignore the weights to then use an unweighted regression. [2] Let $\hat{\epsilon}_u = Y - X\hat{\beta}_u$. Firstly, Pfeffermann & Sverchkov (1999) considered the null hypotheses

$$H_{0,k} : \text{Corr}(\hat{\epsilon}_u^k, W) = 0, k = 1, 2, \ldots$$

For a given k, the sample correlation after the Fisher transformation follows a Normal distribution asymptotically. While k was not bounded, Pfeffermann & Sverchkov (1999) note that the first 2-3 correlations are sufficient for testing the null hypothesis. [10]

Additionally, they proposed regressing W on $\hat{\epsilon}_u^k$ such that

$$E(W \mid \hat{\epsilon}_u^k) = \alpha + \beta^{(k)}\hat{\epsilon}_u^k, k \in \{1, 2, 3\},$$

with intercept α and slope coefficient $\beta^{(k)}$. Afterwards, for a given k, conduct a t-test with $H_{0,k} : \beta^{(k)} = 0$. Lastly, Pfeffermann & Sverchkov (1999) reported the two WA test variations to have similar performance. [10]

3.2.3 Pfeffermann-Sverchkov WA Test (2007)

Pfeffermann & Sverchkov propose another WA test based on regressing W on both X and Y such that

$$E(W \mid X, Y) = \eta X + \gamma Y.$$

By conducting a t-test for the null hypothesis $H_0 : \gamma = 0$, it determines whether the weight is informative for Y which rejecting indicates so. [12] Note that the test was created in the context of small area estimation which Bollen *et al.* (2016) presented it as a more general test for weights. [2]

3.2.4 Wang-Wang-Yan WA Tests

In Wang *et al.* (2023), they collected and simulated several survey weight diagnostic tests. In their review, they propose several modifications to Pfeffermann & Sverchkov's WA tests to account for several existing limitations. [15]

1. **Pfeffermann & Sverchkov (1999)**. Wang *et al.* (2023) notes two notable limitations: multiple testing issues for $k \in \{1, 2, 3\}$ and the regression model for W does not condition on X which may harbor high correlation between W and $\hat{\epsilon}_u$ due to X. They propose a

simple modification by regressing W on the first two moments and an interaction with X:

$$E(W \mid \hat{\epsilon}_u) = f(X; \eta) + \sum_{k=1}^{2} \beta^{(k)} \hat{\epsilon}_u^k + \text{diag}(\hat{\epsilon}_u) X \gamma,$$

where $f(X; \eta)$ is a function of X with scalar parameter η, scalar coefficients $\beta^{(1)}$ and $\beta^{(2)}$, and γ is a $p \times 1$ coefficient vector for the interaction between X and $\hat{\epsilon}$. Finally, test the null hypothesis $H_0 : \beta^{(1)} = \beta^{(2)} = \gamma = 0$ by an F-test. [15]

2. **Pfeffermann & Sverchkov (2007)**. Wang *et al.* (2023) critiques the regression model $E(W \mid X, Y)$ since it would only capture a linear relationship between W and (X, Y). Thus, they suggest capturing possible non-linear relationships by considering

$$E(W \mid X, Y) = f(X; \eta) + \sum_{k=1}^{2} Y^k \gamma_k,$$

where $f(X; \eta)$ is a function of X with parameter η, coefficient γ_k of Y^k. Finally, test the null hypothesis $H_0 : \gamma_1 = \gamma_2 = 0$ with an F-test to determine whether W and Y are associated conditional on X. [15]

3.2.5 Wu-Fuller WA Test

As another special case of the Hausman (1978) misspecification regression test, Wu & Fuller (2005) extend Dumouchel & Duncan (1983) model but change how X is transformed in the regression. Consider the regression

$$Y = X\beta + \widetilde{X}\widetilde{\beta} + \widetilde{\varepsilon},$$

where $\widetilde{X} = QX$, $Q = \text{diag}(q_1, q_2, \ldots, q_n)$, and $q_i = w_i \hat{w}^{-1}(x_i)$ where $\hat{w}(x_i)$ is estimated from the regression of w_i on $f(x_i)$. Testing the model with the null hypothesis $H_0 : \gamma = 0$ determines the impact of W on Y after removing the information contained in X as q_i are the predictable factors of weight W_i by X_i. [16]

3.3 Other Tests

Beyond the parametric WA and DC tests collected and reviewed by Bollen *et al.* (2016) exist a variety of diagnostic tools that may help researchers determine whether weights are necessary in their regression analysis. Some still consist of formal parametric tests or informal judgement calls. Some informal tests or known opportunities to translate fields of thought into tests are noted below:

1. **Bayesian statistics**. Bayesian statistics provides another perspective on weighting, yet there are no proposed tests for weights from a Bayesian perspective despite its potential usefulness as a departure from frequentist statistics. Bayesian inference using linear regressions is an active part of survey data inference literature and available for researchers via the `rstanarm` R-package. See Si *et al.* (2020) for more information. [13]

2. **Standard Errors**. Gelman (2007) notes how standard errors are influenced by the survey design and consider how weighted regressions generally increase standard error estimates. He provides discussion on how to navigate this issue. Thought Gelman (2007) does not offer a diagnostic test, he recommends to use the procedure used to create the weights to compute the standard errors. [5]

3. **Confidence Intervals**. Considered as a DC test by Bollen *et al.* (2016), fitting models with and without weights and assessing whether the associated confidence intervals overlap is a crude diagnostic test when a more formal DC test is not available.

3.3.1 Pfeffermann-Sverchkov Estimation Test

Pfeffermann & Sverchkov (2003) proposed a test that uses the estimating equations to estimate β by an auxiliary regression model for W on some function of X with parameter η. The unweighted estimating function

$$\delta_i(\beta) = X_i(Y_i - X_i^T\beta), i \in S.$$

Define $\hat{W}_i$ as the fitted value of the regression, $q_i = \frac{W_i}{\hat{W}_i}$, and $R(X_i; \beta) = \delta_i(\beta) - q_i\delta_i(\beta)$. Thus, the null hypothesis is $H_0 : E(R(X_i; \beta)) = 0$. The sampling weight means $E(R(X_i; \beta))$ can be tested by a Hotelling statistic

$$\frac{n-p}{p}\bar{R}_n^{-T}\hat{\Sigma}_{R,n}^{-1}\bar{R}_n,$$

where $\bar{R}_n$ is the sample mean and $\hat{\Sigma}_{R,n}$ is the sample variance matrix of $R(X_i; \hat{\beta}_u)$ with $i \in S$. The statistic then approximately follows an F distribution with $(p, n-p)$ degrees of freedom under the null hypothesis. [11]

3.3.2 Breidt Likelihood-Ratio Test

Breidt *et al.* (2013) proposed an LR test which does not fall under a DC nor a WA test. Assuming a superpopulation model with a finite population U, Breidt *et al.* (2013) proposes a weighted log-likelihood with a general weight vector $\omega = (\omega_1, \ldots, \omega_n)^T$ is

$$l(\theta; \omega) = \sum_{i=1}^{n} \omega_i log(f(Y_i \mid X_i; \theta)).$$

Let $\hat{\theta}_U = argmin_\theta l(\theta; U)$ and $\hat{\theta}_W = argmin_\theta l(\theta; W)$. Two LR statistics are considered as

$$T_U = 2(l(\hat{\theta}_U; U) - l(\hat{\theta}_W; U)) \text{ and } T_W = 2(l(\hat{\theta}_W; U) - l(\hat{\theta}_W; W)).$$

Implementing the LR tests require maximizing both weighted and unweighted log-likelihoods. It is important to note that the limiting distribution is a linear combination of chi-square random variables with coefficients being the eighenvalues of the matrix. [3]

4 Existing Simulation Studies

Existing simulation studies include the simulations in the articles that proposed survey weight diagnostic tests and a general review by Wang *et al.* (2023) recently. For the simulation studies alongside proposed diagnostic tests, the simulations are limited in scope to only how the specific test performs without any meaningful comparisons to its characteristics with other tests. Currently, Wang *et al.* (2023) is the sole article that has simulated how the various tests performed in different scenarios comparatively.

The key result is that the test of Pfeffermann & Sverchkov (2007) is the most successful in retaining its size and power among the purely simulated data and the Chinese household consumption expenditure dataset. They also note that most tests were robust to the distribution of the regression error except for the likelihood-ratio test. [15] The limitations of the review is the simulation only considered simple linear regressions as literature does not typically address how the diagnostic tests change when applied to generalized linear models. Additionally, Wang *et al.* (2023) only utilizes simple unequal sampling probabilities without consideration of more complex sampling methods like commonly-used stratified and clustered random sampling in survey data. Sadly, Wang *et al.* (2023) failed to provide the simulation code for reproducibility.

References

[1] Asparouhov, T., and Muthen, B. Testing for informative weights and weights trimming in multivariate modeling with survey data. No. 2, pp. 3394–99.

[2] Bollen, K. A., Biemer, P. P., Karr, F. A., Tueller, S., and Berzofsky, M. E. Are survey weights needed? a review of diagnostic tests in regression analysis. *Annual Review of Statistics and Its Applications*, 3 (2016), 375–392.

[3] Breidt, F. J., Opsomer, J. D., Herndon, W., Cao, R., and Francisco-Fern, M. Testing for informativeness in analytic inference from complex surveys. In *Proceedings 59th isi world statistics congress* (Hong Kong, 2013), pp. 889–893.

[4] DuMouchel, W. H., and Duncan, G. J. Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association 78* (1983), 535–543.

[5] Gelman, A. Struggles with survey weighting and regression modeling. *Statistical Science 22*, 2 (2007), 153–164.

[6] Hausman, J. Specification tests in econometrics. *Econometrica 46*, 6 (1978), 1251–1271.

[7] Kish, L., and Frankel, M. R. Inference from complex samples. *Journal of the Royal Statistical Society 36*, 1 (1974), 1–37.

[8] Lohr, S. L. *Sampling: Design and Analysis*, 3 ed. CRC Press, Boca Raton, 2022.

[9] Pfeffermann, D. The role of sampling weights when modeling survey data. *International Statistical Review 61*, 2 (1993), 317–337.

[10] Pfeffermann, D., and Sverchkov, M. Parametric and semi-parametric estimation of regression models fitted to survey data. *Indian Statistical Institute 61*, 1 (1999), 166–186.

[11] Pfeffermann, D., and Sverchkov, M. *Fitting generalized linear models under informative sampling.* John Wiley Sons, Ltd, Chichester, UK, 2003, ch. 12, pp. 175–195.

[12] Pfeffermann, D., and Sverchkov, M. Small area estimation under informative probability sampling of areas and within the selected areas. *Journal of the American Statistical Association 102*, 480 (2007), 1427–1439.

[13] Si, Y., Trangucci, R., Gabry, J. S., and Gelman, A. Bayesian hierarchical weighting adjustment and survey inference. *Survey Methodology 46*, 2 (2020), 181–214.

[14] Särndal, C.-E., Swensson, B., and Wretman, J. *Model assisted survey sampling.* New York: Springer-Verlag, 1992.

[15] Wang, F., Wang, H., and Jun, Y. Diagnostic tests for the necessity of weight in regression with survey data. *International Statistical Review 91*, 1 (2023), 55–71.

[16] Wu, Y., and Fuller, W. A. Preliminary testing procedures for regression with survey samples. In *Proceedings of the joint statistical meetings, survey research methods section* (2005), pp. 3683–88.