

Continuous Survey Sample Optimization Using Ad Platform APIs

December 11, 2023

Abstract

By reaching billions of people everyday, digital Ad platforms such as Google, Facebook and Instagram offer researchers and advertisers unprecedented opportunities to gather survey responses and conduct experiments entirely online. Researchers have successfully gathered “convenience samples” by creating ads for targetable populations. However, by manually creating these ads, researchers are limited to the explicit targeting criteria available on the platform as well as limited in the complexity of stratification which is feasible. We formulate the problem of dynamically setting ad budgets for survey sample collection as one of minimizing the variance of a post-stratification weighted expectation subject to budget constraints. We use this technique to replicate some questions from important social science surveys, including the General Social Survey (GSS) and the Current Population Survey (CPS). We analyze the differences between our survey responses and those of the major social surveys to show the potential accuracy of our technique. Finally, we present the software as an open-source platform that can be deployed as a web application on public or private clouds.

1 Introduction

The popularization of the telephone in households in the mid-twentieth century led to a rise in telephone surveys as a research method. Initial techniques started by sampling from telephone directories, then developed random digital dialing (RDD) to overcome biases related to unlisted numbers, followed by RDD optimization and improvement techniques to make the process more efficient and cost effective (Cooper 1964; Glasser and Metzger 1972; Sudman 1973).

The twenty-first century, however, has seen steadily rising nonresponse rates for RDD surveys in recent decades, reaching 91% in the 2010s (Keeter et al. 2017; Shirani-Mehr et al. 2018). This level of nonresponse introduces a potential bias that is significantly larger in scale than that of the original bias related to unlisted numbers. What more, phone numbers do not come with demographic variables that allow for the creation of a nonresponse model. To overcome this, it has become standard practice to gather demographic variables in the survey and re-weight responses by comparing to known demographic distributions in the population of interest. This is known as poststratification weighting (Gelman and Little 1997).

The rise of the internet, internet access, and the use of advertising in digital spaces has led to another channel through which organizations can reach individuals: ads. While many researchers are already using ads to recruit sample into their surveys, techniques to recruit efficiently and minimize known biases are still limited.

The purpose of this paper is to posit that:

1. If poststratification techniques are anyway to be used, then support is more important than bias in the initial sampling frame.
2. Digital ads offer strong support across a large portion of the population.
3. Digital ads can be targeted and optimized to increase sample from underrepresented groups to increase the efficiency of the recruitment process.
4. Digital ad platforms offer automated programming interfaces (APIs) that allow for the creation of open-source software that encodes these targeting and optimization techniques and makes them easily available to many researchers.

In what follows, we will make the case for these points, define an optimization technique we have developed, share results from using this technique, and provide access to an open-source software tool that implements these techniques.

2 Non-probability Sampling and Poststratification

The traditional approach of recruiting survey participants starts with probability sampling: the entire population of interest is in the sampling frame and N individuals are selected with known probabilities. This can be contrasted with non-probability sampling, where the selection probability of individuals is unknown to the researcher. Recruiting via digital advertising is a non-probability sampling technique because the researcher does not know the probability that any individual will be shown an ad.

Even in probability sampling, and even if the initial frame is perfect and selection probabilities are known, non-random nonresponse implies that survey responses are not

themselves representative and thus not unbiased out-of-the-box. For this reason it can be helpful to decompose survey error into frame error and nonresponse error. Both, however, create the same problem: your sample is biased and responses must be reweighted in order to create an unbiased estimate of a population parameter (Kolenikov 2016).

Reweighting methods can be separated into three categories:

1. Design weights. If the selection probability of each individual in the sampling frame is known, responses can be weighted by the inverse probability of selection. No external data or modeling assumptions are needed.
2. Nonresponse weights. An estimated probability of nonresponse can be calculated within the survey, given the observed variables recorded in the survey. No external data is required, but the researcher must make modeling assumptions to estimate the nonresponse probability as a function of observable covariates available in the initial sampling frame.
3. Poststratification weights. If external data on the population is available (e.g., census data), response weights for a subpopulation can simply be set to reflect their total proportion of the population. Techniques such as raking¹ can be used if only marginal distributions are available in the population (Deville and Särndal 1992; Battaglia, Hoaglin, and Frankel 2009) and techniques such as multilevel regression and poststratification (MRP)² can be used if the number of variables, and hence subpopulations, is large (Gelman and Little 1997). These techniques do not require known selection probabilities, only non-zero representation of all subpopulations of interest in the final survey results (and by extension, in the original sampling frame). They do require modeling assumptions over the observable covariates available both in the survey data and in the external data.

Design weights together with nonresponse weights are sufficient to recover an estimate of the parameter of interest in the population that is the sampling frame. Alternatively, poststratification weights with external data are sufficient to estimate a parameter of interest in the population represented by the external data. If the initial sampling frame

¹Raking, also known as iterative proportional fitting, consists of iteratively looping over each variable and adjusting the weights of individual observations such that the weighted marginal probability of the variable in the sample approximates that of the population. The process goes like this: given a set of variables, start with one of them and reweight your observations so that the weighted distribution of your sample matches the marginal distribution of this variable in your target population. Repeat for the next variable, and then the next, all the way down the line for all your variables. By the time you adjust for the last variable, your first one is likely off again, so you iterate the process until convergence. While the technique provides no theoretical guarantee on the closeness of the joint density, it has been shown to work well in practice.

²Multilevel regression and poststratification consists of using a Bayesian hierarchical model which has the advantage of being able to estimate cell-specific parameters for populations stratified on many variables, even when the sample size in individual cells is small or non-existent. For example, consider a survey in the US where you would like to include the variables race, gender, and state. You might not have many observations of each combination of race/gender within each state (or even just many observations within each state period). However, if you model states as belonging to higher-level groups (regions), you can estimate the region-level parameters and then regularize the state-level parameters to the corresponding regional-level parameters. This allows the individual state-level parameters to deviate from the region only when there is enough state-level data to justify the deviation. Given estimates for every cell in your stratification, poststratification weighting consists in directly applying the target population weights for the cell.

is the same as the "external data" about your target population (i.e. both are an official government census), then the two techniques are identical.

In practice, however, many surveys start with a sampling frame that is only somewhat representative and not very covariate-rich. This is the case with random-digit-dialing (RDD), the most common technique used by companies and organizations conducting phone surveys and a staple of opinion polling for many years. Due to rising nonresponse in these surveys, it has become standard practice for these surveys to employ poststratification weighting to estimate their population parameter (Gelman and Little 1997).

What, then, is the value of a quasi-representative initial frame if nonresponse is so significant and poststratification techniques needed anyways? Can nonprobability sampling techniques with poststratification replace traditional probability-based sampling measures? If so, in what settings is this feasible and advisable?

These questions have been posed by a number of researchers of late. Wang et al. (2015) applied multilevel regression and poststratification (MRP) to data from an opt-in poll that was made available on the Xbox gaming platform to estimate vote share among the two primary candidates for the 2012 US presidential elections. They compare their predictions with polling averages from pollster.com and find that their predictions track the polling averages very closely and indeed even produce better results than the polls in the days running up to the election. Goel, Obeng, and Rothschild (2015) applied MRP to survey data from a sample of 1000 Amazon Mechanical Turk (AMT) workers and from 1,000 respondents collected via online survey company Pollfish to calculate mean outcomes for a variety of questions from the General Social Survey (GSS) and similar questions from surveys performed by Pew Research. They report that their calculated outcomes from AMT and Pollfish respondents have a mean-absolute deviation (MAD) from GSS/Pew benchmarks of 7.2 and 7.4 percentage points respectively, where the GSS and Pew differ from each other with a MAD of 8.6. Further research from Pew shows a similar error magnitude between online opt-in surveys with poststratification methods and their traditional phone-based survey results, reporting average deviations of 6 percentage points (Mercer, Lau, and Kennedy 2018).

What do these results imply? The GSS is a rigorous survey in which considerable time and resources are poured into creating an inclusive sampling frame and then getting responses (in-person) from each sampled individual (minimizing nonresponse), coming in at the considerable price tag of \$3 per question per respondent (Goel, Obeng, and Rothschild 2015). While there is no ground truth in the opinion questions measured in these surveys, the GSS is widely considered the best we have. Phone-based surveys from firms like Pew Research start with a significantly more representative sampling frame (all phone numbers) than the convenience samples studied (MTurkers or opt-in visitors to a set of websites/apps). Despite this advantage, it does not systematically outperform the convenience sample. This would seem to imply that either A) the magnitude of nonresponse error in those techniques overwhelms any improvement in frame error and/or B) that poststratification techniques successfully made up for a significant portion of frame error, reducing the importance of the initial frame in the resulting total survey error.

Critically, however, poststratification techniques (and nonresponse modeling) require the researcher to select a set of relevant variables on which to stratify. This need to model the outcome as a function of covariates is happily absent in a pure probability-sampling method with minimal nonresponse.

A simple example can help illustrate this point: imagine you are interested in surveying your city's population to know how concerned they are about a particular global pandemic.

Consider, additionally, that you collect information on respondents’ race and gender, but do not think to collect information on their age. Assume that in reality, older people are more vulnerable to this disease and thus more concerned.

If you have a comprehensive sampling frame and a crack team of door-knockers (think GSS), your resulting sample (given sufficient size) will likely be representative for age and thus your lack of information about respondents’ age will not affect the estimate of the average level of concern in your city. Now consider you are running an RDD-based phone survey and it turns out that young people are much less likely to take your call. You did not collect information on age (you did not know it would matter!), so poststratification cannot make up for this nonresponse bias and you end up woefully overestimating the concern in your city. Conversely, consider you run an internet-based convenience survey in which young people are overrepresented in your sampling frame. Without information on the age of each respondent in your collected data, poststratification cannot overcome this bias.

While the importance of “age” might be obvious in modeling citizens’ concern regarding a pandemic, the inclusion or exclusion of other variables might not be so obvious a priori (e.g., political affiliation or social media diet). Thus, any technique relying on poststratification implicitly brings a modeling and, in particular, a variable-selection problem. Indeed, Mercer, Lau, and Kennedy (2018) show in a comparison of poststratification techniques that what matters most for removing the bias of nonprobability samples is not the exact technique used, but rather the variables chosen. This should be a strong cautionary tale to anyone who thinks they have a representative sample but has taken the variables required for representation as given and not chosen them based on the specific outcome. In the following section we will discuss how integrating recruitment and survey responses can allow for the variable selection problem to be formally solved in an online fashion, potentially leading to significant increases in outcome estimation accuracy.

An additional disadvantage of a non-inclusive sampling frame is that, while poststratification weighting can make up for certain populations being underrepresented, it can do nothing for populations entirely absent from the sampling frame. Digital advertising, along with the other online convenience methods compared in the studies reported above, by definition exclude from their initial sampling frame everyone without internet access. This could potentially exclude very poor households entirely, as well as those of certain religious or ethical beliefs. This is a deficit that no amount of poststratification weighting can make up for.

Many new studies have begun to use Facebook specifically as a recruitment tool and apply poststratification to estimate population quantities (Zagheni, Weber, and Gum-madi 2017; Perrotta et al. 2020). We are not aware, however, of any study that has systematically compared recruitment via online advertising to traditional probability-based recruitment such as RDD. We see several advantages that online advertising has as a sampling frame as opposed to other online convenience sampling methods:

1. Large population coverage: Facebook registers 2.7 billion active users (Statista 2020). Google Ads Display Network reports to cover 90% of internet users worldwide across millions of websites (Google 2020).
2. Targeted advertising allows researchers to intentionally pay more to reach under-represented groups. This allows researchers to trade off cost and representativeness in a way that the studied convenience methods (i.e. Xbox live players) do not allow.

3. Real time communication. Digital advertisers expose APIs, which allows software run by the researcher to communicate with the advertising platform in real time and adjust ad placement. While on the one hand this makes it convenient to create hundreds of audiences for the stratified recruitment, the potential extends further. We explore this feature more in the following section, but it's worth highlighting that it is novel, not present in traditional sampling frames or considered in any of the research on this topic that we are aware of.

The combination of the last two points (targets ads controlled via API) is very powerful: it implies that we can build our own ad-optimization engine to optimize the goals of researchers, policymakers, and the public good. This does not come out-of-the-box from ad platforms, whose built-in ad optimization routines are designed to maximize value under the assumption that diversity of audience (customers) is not in-and-of-itself valuable. While this may be the case in retail, it is not the case for research. The value (information gain) of an individual decreases with the number of similar individuals we already have. This is why Virtual Lab has its own ad optimization engine that uses the available tools to optimize for heterogeneity rather than homogeneity.

These additional advantages of digital advertising over the convenience-based methods studied give solid reasons to believe that results could potentially be even better for digital advertising. Future research is needed to test that hypothesis.

3 An Online Algorithm for Optimizing Stratified Recruitment

We consider that the researcher is starting with a well-defined population of interest P and would like to estimate a population parameter Y via sampling and measurement.

We will consider the case, without loss of generality, where measurement is performed by administering a survey to each sampled individual and the outcome is a survey response. We will consider that the survey might be divided into one or more waves administered at different points in time.

Sampling and administering the first wave of the survey will be performed by a process jointly referred to as “recruitment”. We consider that the sampling frame to be the set of individuals that can be targeted and reached directly via a digital ad platform, denoted F . For this technique to work, we assume that $F \cap P$.

We denote an individual outcome response as y_i for each individual i in the sample. We do not specify the survey wave in which the outcome response is recorded. We will assume that the researcher wishes to use, as an estimator of Y , the stratified mean for a set of strata $h \in H$ with an assigned weight for each stratum W_h , which we will denote \bar{y} :

$$\bar{y} := \sum_h W_h \bar{y}_h$$

where \bar{y}_h denotes the sample mean within stratum h . We do not specify the motivation for the stratification or the weights. We do assume that we are able to measure a set of additional survey responses which we will consider covariates and denote $x_i \in X$. We assume the existence of a mapping $X \rightarrow H$ such that the measured covariates are sufficient to assign each individual to one and only one stratum. In addition, we assume that x_i is measured during recruitment.

The variance of our sample estimate is thus given by:

$$\mathbb{V}[\bar{y}] = \sum_h W_h^2 \frac{s_h^2}{n_h}$$

where s_h^2 denotes the variance of the population parameter of interest Y within stratum h . If the outcome was measured during recruitment, we can estimate this stratum-specific variance. We will do this in the sequel (section to be written), however, we will begin with a simplification that makes this sampling procedure more generally useful by including scenarios where the outcome response is not measured during recruitment but at a later time. We simplify the problem by assuming that the variance of the outcome in each stratum is equal (i.e. $s_h^2 = s^2$). With that assumption, we have the following variance of our estimate:

$$\mathbb{V}[\bar{y}] = s^2 \sum_h \frac{W_h^2}{n_h}$$

Note that, given a fixed n and the assumption of equal variance across strata, this quantity is minimized when $\frac{n_h}{n} = W_h$, known as the Neyman allocation (Neyman 1934; Groves et al. 2010). If there were no friction in recruiting individuals in the sampling frame from each stratum, then the researcher would simply pick the Neyman allocation. However, in the real world, there is friction in sampling individuals from different strata and, furthermore, that friction might differ greatly across strata.

In particular, we will assume that this friction is represented as a monetary cost and denote the cost to recruit an individual from stratum h as P_h . We then assume the researcher has a fixed budget B to allocate to the recruitment of individuals in each stratum such that $B_h = P_h n_h$. We also assume that the researcher has a desired maximum sample size that they can measure, which we will denote N_d . We can then frame the optimization problem of finding the best allocation of budget to minimize the variance of the final estimate as:

$$\begin{aligned} \operatorname{argmin}_{n_1, \dots, n_h} \quad & \sum_h \frac{W_h^2}{n_h} \\ \text{s.t.} \quad & \sum_h P_h n_h \leq B \\ & \sum_h n_h \leq N_d \end{aligned}$$

Note that this problem will not necessarily have a unique solution. However, given that the variance is monotonically decreasing for each n_h and the constraints both cap the maximum possible n_h , any minimum will result in at least one non-slack constraint. This can be used to simplify the computational search.

If P_h was known, then this optimization problem would allow us to determine the optimal spend per stratum B_h and that spend could be allocated. However, in many real-world scenarios, P_h may not be known ahead of time.

We thus propose a simple algorithm to continuously both estimate the price and optimize recruitment. To do so, we require an interface for recruitment that targets stratum h and allocates budget B_{ht} over a specific period of time t . It should be noted that

digital ad platforms provide just such an interface. We denote this interface $\text{Recruit}(B_t)$ which accepts a budget allocation $B_t := \{B_{1t}, \dots, B_{Ht}\}$.

Additionally, we require an interface $\text{GetResults}(t)$ to collect information on the results of recruitment at time t given budget B_t . Results should be considered as the number of respondents recruited for each stratum h at time t and will be denoted n_{ht} .

Given that, we can model the inverse cost ($\frac{1}{P_h}$), the number of respondents recruited n_{ht} given budget spend B_h , as a Poisson random variable:

$$\begin{aligned} n_{ht} | B_{ht} &\sim \text{Poisson}(\lambda) \\ \lambda &\sim \text{Gamma}(\kappa, \beta) \end{aligned}$$

If we set a prior for the parameter λ in the form of parameters of a Gamma distribution (κ and β), we can use closed-form Bayesian updating to obtain a MAP estimator of λ and the implied mean of the predictive distribution ($1/\lambda$). The estimated mean number of respondents per unit of budget can then be used to calculate the price which will in turn be used for the optimization.

The problem with this method is, of course, that we are not taking into account the uncertainty in our price estimate when we solve our optimization problem. Additionally, here we are assuming that there is a single, static price for each stratum. However, in the case of digital ad spend, our price is actually a function of the budget allocated. We leave it as a task for future work to fully model the whole system.

Given these simplifying assumptions, however, we can build an extremely simple and flexible algorithm to optimize ads for recruitment:

Algorithm 1 Optimizing Stratified Recruitment with Unknown Costs

```

procedure ESTIMATEPRICE( $\kappa, \theta, B_t, n_t$ )      ▷ Priors  $\kappa, \theta$  set from domain knowledge
     $\beta := 1/\theta$ 
     $\lambda := (\kappa + n_t)/(\beta + B_t)$ 
     $p := 1/\lambda$ 
    return  $p$ 
end procedure
procedure ADOPTIMIZATION( $W, B, N_d, \kappa, \theta$ )
     $B_0 := [1, \dots, 1]$                                 ▷ Budget indexed by H strata
     $n := [0, \dots, 0]$                                     ▷ Results indexed by H strata
    for  $t \in T$  do
         $p_t := []$                                           ▷ Price estimates indexed by H strata
        for  $h \in H$  do
             $n_{ht} := \text{GetResults}(h, t)$ 
             $n_h := n_h + n_{ht}$ 
             $p_{ht} := \text{EstimatePrice}(\kappa, \theta, B_{ht}, n_{ht})$ 
        end for
         $B_{t+1} := \text{Optimize}(W, B, N_d, n, p_t)$ 
         $\text{Recruit}(B_{t+1})$ 
    end for
end procedure

```

4 Connecting the Algorithm to Ad Platform and Survey Platform APIs

Digital ad platforms are, by definition, software. Humans interact with software through interfaces. Often, this is a “graphical user interface” (GUI), such as when an advertiser creates ads directly on the ad platforms website.

Most modern ad platforms additionally offer an automated programming interface (API) which allow software to interact with the platform and perform all the actions that a human can perform via the GUI.

The existence of such APIs implies that the Recruit interface required by the AdOptimization procedure described previously can be implemented as procedures that consume the APIs that the ad platform publishes.

What, then, about the GetResults interface? If the results are stored in an internet-connected software platform that also publishes an API, and allows the retrieval of the results via such an API, then GetResults, too, could be implemented in the same fashion. Imagine you are using a web survey platform such as Qualtrics, Typeform, SurveyMonkey, or Alchemer. All of these platforms publish an API that allow us to programatically collect survey responses, complete with information about the respondent and the time of the response. Thus, at any given time, we can know how many people have completed the survey. Given our assumption about the existence of a mapping between $X \rightarrow H$, we can also know the stratum of each survey taker, allowing us to calculate in real time how many people have completed the survey in each stratum.

Note that survey platforms are not the only platforms in which we might want results and are not the only platforms in which we can collect results via API. For example, one can imagine that instead of directing respondents to a survey, one wishes them to download and install an app. As long as the usage data of that app is available via an API, the same technique can be used.

The result is that the entire AdOptimization procedure can be written as software that runs autonomously and re-optimizes the recruitment ads as often as required (e.g. every four hours). We have written software that does this: the sourcecode is open and can be found in our public git repository³.

5 Results

We have run 35 studies in 24 countries, with recruitment ads reaching 33,073,461 people and 166,535 final respondents optimized with this technique. We have created an open dataset with the recruitment information. We include below a table with some noteworthy highlights for each study.

1. **Strata:** the number of different strata used for recruitment
2. **Max CTR:** represents the maximum unique click through rate of any stratum within the study.
3. **Reach:** The total number of people reached by the recruiting ads.
4. **Responents:** The total number of people who started, qualified, and finished the survey. Note this will depend on qualification for the exact survey.

³<https://github.com/vlab-research/vlab>

5. **Average Cost** The average ad spend cost per recruited respondent who qualified and completed the survey. Note that this does not include any advertised incentive, it only includes the advertising spend.

	Country	Strata	Max CTR	Reach	Respondents	Average Cost
0	India	24	5.09%	873653	9130	\$0.24
1	Libya	176	27.1%	1000294	8338	\$0.3
2	Labanon	48	15.03%	1370234	17399	\$0.57
3	Jordan	192	10.05%	2223793	17223	\$0.6
4	Iraq	144	8.58%	4280255	16015	\$0.61
5	Serbia	48	13.58%	985109	13995	\$0.67
6	Nigeria	23	6.42%	145437	2393	\$0.75
7	US	10	13.61%	16849	1118	\$0.85
8	US	4	6.9%	10616	316	\$0.87
9	Haiti	16	10.0%	1218842	10491	\$0.94
10	Honduras	144	7.4%	909597	4922	\$0.95
11	Lebanon	48	13.82%	1650052	14354	\$1.03
12	Papa New Guinea	64	8.71%	118980	1825	\$1.46
13	Iraq	80	8.27%	5616663	6520	\$1.55
14	US	14	6.14%	120647	2553	\$1.66
15	Ukraine	64	5.17%	648512	2394	\$1.69
16	Kyrgyzstan	24	9.63%	489054	3004	\$1.76
17	Djibouti	16	10.84%	313493	2252	\$2.19
18	Kosovo	56	19.93%	663059	6084	\$2.46
19	Chad	32	7.59%	327048	2305	\$2.64
20	Jamaica	16	15.03%	482097	4105	\$2.72
21	Belize	24	9.1%	43684	264	\$2.89
22	Serbia	1	6.75%	342403	1737	\$2.99
23	Macedonia	32	24.58%	384956	3156	\$3.19
24	Romania	200	9.31%	785811	1863	\$3.23
25	Macedonia	32	25.17%	538295	4565	\$3.26
26	Jordan	96	7.14%	1304979	2146	\$3.72
27	US	16	3.93%	241134	1429	\$4.55
28	Nigeria	120	7.75%	563939	177	\$5.55
29	Bulgaria	8	3.91%	170205	170	\$6.16
30	Cameroon	80	11.85%	1427793	1712	\$6.72
31	India	160	5.77%	3526970	1639	\$7.85
32	Gambia	2	5.86%	279008	697	\$11.07

6 Discussion

There are a few things that we believe our results in the last couple of years show quite clearly:

1. It is possible to reach and recruit highly-stratified samples in many countries via digital ad campaigns in a cost-effective manner.
2. A non-trivial percentage of people reached can be inticed to click on the ad.

3. Survey recruitment can be optimized in real-time by building software that connects with ad platform and survey platform APIs.

At the same time, we believe there is still a lot of work to be done. For one, the optimization can be improved to learn the price and better estimate the price function in order to better estimate the cost and time needed to recruit. Time can be an important factor, the slower recruitment goes the more cost effective it is, however it is not trivial to estimate the right speed for a given study. Similarly, we believe that effective ads, building trust with a brand, and strong incentives can increase response rates dramatically as well.

Additionally, another important piece of outstanding work is to show that poststratification weighting applied to these surveys delivers results that compare with gold-standard population surveys and that optimizing for post-stratification weighting during recruitment leads to better results than naive recruitment without stratification.

References

- [1] Michael P. Battaglia, David C. Hoaglin, and Martin R. Frankel. “Practical Considerations in Raking Survey Data”. In: *Survey Practice* 2.5 (2009), pp. 1–10. ISSN: 2168-0094. DOI: 10.29115/sp-2009-0019.
- [2] Sanford L Cooper. “Random Sampling by Telephone - An Improved Method”. In: *Journal of Marketing Research* 1.4 (1964), pp. 45–48.
- [3] Jean-Claude Deville and Carl-Erik Särndal. “in Survey Sampling Calibration Estimators”. In: *Journal of the American Statistical Association* 87.418 (1992), pp. 376–382. ISSN: 01621459. URL: <http://www.jstor.org/stable/2290268>.
- [4] Andrew Gelman and Thomas C. Little. “Poststratification Into Many Categories Using Hierarchical Logistic Regression”. In: (1997).
- [5] Gerald J Glasser and Gale D Metzger. “Random-Digit Dialing as a Method of Telephone Sampling”. In: *Journal of Marketing Research* 9.1 (1972), pp. 59–64.
- [6] Sharad Goel, Adam Obeng, and David Rothschild. “Non-Representative Surveys: Fast, Cheap, and Mostly Accurate”. In: (2015), p. 27.
- [7] Robert M. Groves et al. *Survey methodology*. 2010, pp. 21–64. ISBN: 0472098489. DOI: 10.4324/9780429314254-2.
- [8] Scott Keeter et al. “What Low Response Rates Mean for Telephone Surveys”. In: (2017), pp. 1–39. URL: <http://www.pewresearch.org/wp-content/uploads/2017/05/RDD-Non-response-Full-Report.pdf>.
- [9] Stas J. Kolenikov. “Post-stratification or a non-response adjustment?” In: *Survey Practice* 9.3 (Aug. 2016), pp. 1–12. ISSN: 2168-0094. DOI: 10.29115/SP-2016-0014. URL: <https://surveypractice.scholasticahq.com/article/2809-post-stratification-or-non-response-adjustment>.
- [10] Andrew Mercer, Arnold Lau, and Courtney Kennedy. “For Weighting Online Opt-In Samples, What Matters Most?” In: *Pew Research Center* (2018).
- [11] Jerzy Neyman. “On the Two Different Aspects of the Representative Method : The Method of Stratified Sampling and the Method of Purposive Selection Author (s) : Jerzy Neyman Source : Journal of the Royal Statistical Society , Vol . 97 , No . 4 (1934) , pp . 558-625 Pub”. In: *Journal of the Royal Statistical Society* 97.4 (1934), pp. 558–625. ISSN: 0952-8385.

- [12] Daniela Perrotta et al. “Behaviors and attitudes in response to the COVID-19 pandemic: Insights from a cross-national Facebook survey”. In: (2020), pp. 1–17. DOI: 10.1101/2020.05.09.20096388.
- [13] Houshmand Shirani-Mehr et al. “Disentangling Bias and Variance in Election Polls”. In: *Journal of the American Statistical Association* 113.522 (2018), pp. 607–614. ISSN: 1537274X. DOI: 10.1080/01621459.2018.1448823. URL: <https://doi.org/10.1080/01621459.2018.1448823>.
- [14] Seymour Sudman. “Uses of Telephone Directories for Survey Sampling”. In: *Journal of Marketing Research* 10.2 (1973), pp. 204–207.
- [15] Wei Wang et al. “Forecasting elections with non-representative polls”. In: *International Journal of Forecasting* 31.3 (2015), pp. 980–991. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2014.06.001. URL: <http://dx.doi.org/10.1016/j.ijforecast.2014.06.001>.
- [16] Emilio Zagheni, Ingmar Weber, and Krishna Gummadi. “Leveraging Facebook’s Advertising Platform to Monitor Stocks of Migrants”. In: *Population and Development Review* 43.4 (2017), pp. 721–734. ISSN: 17284457. DOI: 10.1111/padr.12102.