Survey Weighting with Differentially Private Releases of Population Data

An Evaluation of the Cooperative Congressional Election Study

Bhaven Patel and Anthony Rentsch

Spring 2019

Abstract

Beginning in 2020, the American Census Bureau has announced that it will release some of its data products in a differentially private fashion, a change which introduces new considerations for researchers who rely on Census data. In this paper, we focus on the effects of differential privacy on statistical methods that use Census data for calibration; in particular, we look at the case of survey weighting. We find that, for smaller populations, weighting a survey to a differentially private release of the American Community Survey can shift estimates of important survey items by a small margin, but that this shift vanishes when the privacy loss parameter is incremented, i.e., when less noise is added to the released data.

1. Introduction

A key methodological component of survey research is weighting, a technique that tries to correct imbalances between the composition of the sample and the composition of the true population of interest. For a variety of reasons, it is extremely difficult to collect a simple random sample of the American adult population, so survey researchers must gather samples through other methods, which often lead to samples that do not accurately reflect the composition of the population. To obtain credible population-level estimates, it is necessary to re-weight individual respondents in a sample to accurately reflect the population as a whole.

While there are numerous different methods to compute survey weights, there is one common feature across these methods: the reliance on American Census Bureau data to estimate true, benchmark population parameters. In particular, the Census' American Community Survey (ACS) is a high-quality survey that produces estimates of population demographics and socioeconomic conditions at various geographic levels in the United States, and is typically regarded as one of the most reliable population benchmarks.

Beginning in 2020, the Census has announced that it will release some of its data products in a differentially private fashion in order to protect the confidentiality of individuals who participate in the decennial census, as well as in other surveys that the bureau fields. One useful notion of differential privacy is as a mathematical formulation of privacy that provides guarantees about what an adversary is able to learn about any individual in a dataset after performing some statistical analysis to glean information from the dataset. Intuitively, differential privacy adds noise to a statistical or machine learning analysis so that it is difficult to tell whether or not any individual's information was used as a part of the analysis.

Currently, the Census is conducting research regarding transitioning the ACS to formally private methods, such as differential privacy [3]. Regardless of what the final implementation is, the introduction of differential privacy would add new considerations into the work of analysts who use the ACS. There is evidence that the Census' historical disclosure avoidance methods have introduced non-trivial amounts of error into its data products [2]. It is currently unclear how much differential privacy would impact those who work with ACS data and the extent to which these alterations could be accounted for by modeling the mechanism that produced the private statistics, which is a benefit of differential privacy that is not possible with other methods that seek to preserve individual-level privacy.

The focus of this paper is on the potential effect of differential privacy on statistical models that rely on Census data for calibration. Specifically, we analyze the impact that a differentially private release of the ACS

could have on the computation of survey weights for one widely-used, large sample size political survey: the Congressional Cooperative Election Study (CCES). We compare subgroup estimates for various survey items when the survey is weighted to the actual ACS data versus when the survey is weighted to a hypothetical differentially private release of the ACS data. We find that, for smaller populations, weighting the CCES to a differentially private ACS release can shift estimates of important survey items by a small margin, but that this shift vanishes when the privacy-loss parameter is incremented, i.e., when less noise is added to the released data.

To our knowledge, this is not a topic area that has received scholarly attention. One recent study looked at generating survey weighted frequency tables under differential privacy, but that analysis focused more on the effects of adding noise to survey data that had already been weighted rather than adding noise to the benchmark data to which the survey is weighted [5].

We emphasize that this analysis is a first cut at understanding the impact of differential privacy on statistical calibration methods for three reasons: (1) there still remains an incredible amount of uncertainty regarding how the Census Bureau will implement differential privacy for the ACS, (2) calculating the appropriate amount of noise to add to ACS data is challenging, and (3) surveys like the CCES employ incredibly complex sampling and weighting procedures, which are beyond the scope of this project to replicate. Future work will need to take into account all of these considerations more seriously. For now, we think that our stylized example provides a useful look into what new issues consumers of Census data will face with the rise of differential privacy.

2. Key Background

2.1 Differential privacy

Differential privacy is a criteria for a statistical or machine learning mechanism that holds that an adversary should be able to make the same inference about an individual's sensitive information whether or not that individual was included in the analysis [6]. In mathematical notation, this implies that $P[M(x) \in T] \leq e^\epsilon P[M(x') \in T]$ for some mechanism M, two neighboring datasets x and x' for which the attributes of exactly one row have been changed, and some set T. Note that this definition constitutes bounded differential privacy, as the definition of neighboring datasets relies on changing the values of one row rather than adding or removing one row. In this definition, ϵ is a tunable privacy-loss parameter: for smaller values, it implies that the probability distributions of this mechanism acting on neighboring datasets are virtually indistinguishable (meaning more privacy) while for larger values it implies that the probability distributions are easy to discriminate between (meaning less privacy).

While there are many proposed differentially private mechanisms, we rely on one of the most heavily-vetted methods: the Laplace mechanism. This mechanism simply computes a statistic of interest, such as a mean, and adds random noise drawn from the Laplace distribution to that statistic. This mechanism is guaranteed to be differentially private so long as the scale of the random noise drawn from the Laplace distribution is set to the global sensitivity of the mechanism divided by ϵ, where global sensitivity is defined as $max_{x,x'} \mid M(x) - M(x') \mid$ for neighboring datasets x and x' and a mechanism M.

2.2 Census' plan to implement differential privacy

While the Census has announced that it will implement differential privacy for some of its data products starting in 2020, the exact implementantion details are still being developed. Furthermore, the Census' proposed implementation will require a principled way to blend differential privacy and other formal privacy methods. This comes along with other issues, including dealing with summary statistics for which there can be no added noise (frequently called invariants) and imposing structural zeroes for counts of certain populations, like three-year old grandmothers [3].

The ACS poses even more challenges, including working with high dimensional data, generating geographic-specific estimates, preserving household associations, handling outliers in economic variables, and dealing with the ACS' survey weights [3]. At a fundamental level, differential privacy is designed to protect individuals, which has led some to raise concerns regarding the feasibility of generating credible small-area estimates via the ACS [3]. With these challenges in mind, we develop a stylized example to produce reasonable estimates of the expected effects of differential privacy on survey weighting. This is described in more detail in later sections.

2.3 Survey weighting

Since it is difficult for survey researchers to conduct simple random sampling on a large population of interest, they are often forced to obtain samples through other methods and are frequently left with samples that do not reflect the demographics of the population. In order to produce estimates for the population, it is thus desirable for survey researchers to compute weights that give more importance to responses from populations underrepresented in the sample and less importance to responses from populations overrepresented in the sample.

Many methods exist to compute survey weights. For a non-exhaustive list, see this summary compiled by the Pew Research Center [4]. The method that we focus on in this paper is cell weighting, which is a form of post-stratification weighting. For cell weighting, a researcher needs to have the full joint distribution of the population of interest. Using the joint distribution, they are able to analytically compute weights that rebalance the sample to look more like the population with respect to the variables contained in the joint distribution.

Since it requires the full joint distribution of the population, cell weighting is not always feasible in practice. This is less of a concern when the population is U.S. adults since we can estimate the joint distribution using Census data products like the ACS. Furthermore, we think that differential privacy could affect methods like survey weighting by altering small cell counts in the joint distribution. This makes cell weighting an ideal candidate method for this project.

3. Methodology

3.1 Data aggregation and processing

Our survey data comes from the Cooperative Congressional Election Study (CCES) conducted in 2016.[1] The CCES is a large-N national survey conducted in every election year. It samples tens of thousands of respondents in a nationally stratified manner, collecting many of the same demographic variables as the ACS as well as answers to a variety of social and political questions.[2] These questions measure respondents' behaviors, such as how they consume news and how frequently they use social media, and their attitudes regarding important topics such as gun control, abortion, and taxes. To map CCES responses to estimates for the national population, the survey employs a fairly complex sampling and weighting procedure based on population estimates aggregated from the ACS, the Current Population Study (CPS), and the Pew Research Center Religious Landscape Survey. To simplify our task, we decided to only generate survey weights for the CCES respondents based on their demographic data compared to the population estimates provided by the ACS.

To be consistent with the 2016 CCES, we use the 2012 ACS 5-year estimates as our population estimates. In particular, we use the ACS' Public Use Microdata Sample (PUMS) files, which contain a subsample of individual ACS responses.[3] After aggregating the PUMS files for each of the 50 states and Washington D.C., we have over 15 million records. Each record contains a host of demographic and socioeconomic variables

---

[1]https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910/DVN/GDF6Z0
[2]https://cces.gov.harvard.edu/
[3]https://www.census.gov/programs-surveys/acs/technical-documentation/pums.html

about each respondent, as well as a person-level weight, which we will refer to as "person weights." These person weights are computed via a "three-dimensional raking-ratio estimation procedure" that tries to balance benchmarks for (1) the population size of each respondent's subcount, (2) the number of housing units, and (3) the demographic composition of the population [1]. The benchmarks that the ACS uses in this step come from the Census Population Estimates Program, which are derived from the 2010 decennial census [1]. The figure in Appendix A1 shows that the person weights range from 1 to 549, with 99 percent of the weights falling below 77. This data represents our estimate of the full U.S. population. Hereinafter, we refer to this aggregated PUMS data as the ACS estimates.

Before we generate the cell counts for specific combinations of demographics in the ACS, we limit the demographic characteristics to state, race, education, sex, and age. Additionally, we bin the values/categories for race, education, and age in a consistent manner for the ACS and CCES datasets. This method of binning is standard practice among survey researchers, so we chose to take advantage of it here to reduce the number of cell counts in the ACS dataset. For race, every individual in the ACS and CCES is labeled as "White," "Black," "Asian," "Hispanic," or "Other." For education, individuals are labeled as "No HS" (did not graduate from high school), "HS graduate" (graduated from high school), "Some College" (attended some college or university), "2-year degree" (attained an associate's degree), "4-year degree" (attained a bachelor's degree), and "Postgraduate degree" (attained a graduate degree such as a Master's or Ph.D.). Lastly, individuals' ages are recoded into age ranges: less than 35 years old, between 35 and 50, between 50 and 65, and older than 65 years old.

3.2 Cell weighting

We estimate the joint distribution of the U.S. population by computing five-way cell counts in the ACS data for each combination of state, race, education, sex and age based on this bins described above. This produces 12,240 unique combinations among this set of five features. For each bin, the person weights of the individuals in that bin were summed to estimate the total number people in the national population that have this specific combination of demographics. Roughly one percent of cells have counts of zero, indicating that the ACS data contained no individuals with this combination of attributes. We choose to set the cell counts for these bins to one to leave open the possibility that an individual with this combination of features could exist in the population but that it is still extremely unlikely. We use the R *survey* package to calculate the post-stratification weights for each respondent in the CCES based on the demographic cell counts from the ACS data. We use these as the "true weights" for the CCES respondents; the CCES computes its own set of post-stratification weights but we largely ignore these in our analysis. (We ran into some trouble using the R *survey* package to generate weights and a small proportion of our observations were not assigned weights. However, we do not believe that this affects our findings. In the future, it would be beneficial to implement a cell weighting algorithm from scratch to address this and to allow more flexibility in accounting for the noise added).

We also generate a second set of survey weights based on a differentially private release of the ACS cell counts using the Laplace mechanism, which we call "noisy weights" for the CCES respondents. To generate differentially private ACS cell counts, we add Laplace noise to each cell count.

In order for the Laplace mechanism to meet the differential privacy criteria, the scale of the noise added from the Laplace distribution has to be equal to the global sensitivity divided by ϵ. However, determining the sensitivity of the ACS data is not straightforward. Under bounded differential privacy, we define two neighboring datasets x and x' such that an individual is moved into a different state x gender x race x age x education bin. Since we propose adding noise to bin counts, which are the sums of all the weights of people who fall in that bin, a plausible first step might be to set the global sensitivity to 2·max(all person weights). This assumes that moving a person from one bin to another would subtract that person's weight from one bin and add it to another without affecting the counts in any other bin. In the worst case, the weight of this person would be the maximum weight in the dataset; in the 2012 5-year estimates, this value is 549. Under this assumption, we could draw noise from a Laplace distribution centered at 0 and with a scale of $\left(\dfrac{2 \cdot 549}{\epsilon} \right)$.

This reasoning is flawed because it assumes that the ACS person weights are fixed attributes of individuals. In fact, the ACS person weights depend on the weights of other respondents. Changing the bin that one person exists in can actually have a substantial effect on the weights of other respondents. This means that 2·max(all person weights) is not an upper bound on the global sensitivity at all. At this time, it is not entirely clear how to derive the global sensitivity for this mechanism given the ACS weighting methodology. We leave this issue for future works.

Instead, we choose to use 2·max(all person weights) as our sensitivity for the purposes of this analysis. In response to discussions we have had with the Census team working on incorporating differential privacy into the ACS, we consider this value to be a surrogate value for the sensitivity under a differentially private release of the ACS. Under this interpretation, the amount of noise we add to cell counts should not be considered to be an upper bound but rather a reasonable amount given the aims of the Census. Choosing to operate under the bounded differentially private paradigm rather than the unbounded one also allows us to use the scaling factor of 2, which inches our estimate of the amount of noise that will be added closer to the worst case scenario.

After adding noise to the cell counts, we again enforce bin positivity. This operation - $max(1, n_{noisy})$ - is a post processing step, so we do not have to consider allocating any of our privacy budget ϵ here. We do acknowledge that this step may induce a bias during the computation of the post-stratification weights, which we discuss later.

4. Results

We perform several simulations to evaluate the impact of releasing a differentially private version of the 2012 5-year ACS estimates on the computation of post-stratification weights for the 2016 CCES and the subsequent estimates for various survey items. We focus on two survey items: vote preference for the 2016 general election and support for an assault rifle ban. For the vote preference item, our estimand is the difference between the proportion of respondents who say they will vote for Hillary Clinton and the proportion of respondents who say they will vote for Donald Trump (for brevity, this will be called the Clinton lead). For the assault rifle ban item, our estimand is the proportion who support the ban minus the proportion who oppose the ban, or the net support for an assault rifle ban.

Our primary concern is that differentially private releases of population data could introduce relatively large amounts of error into the ACS cell counts for smaller populations and cause the resulting post-stratification weights to be incorrectly calibrated for some demographic subgroups. If this is the case, we would also expect that CCES estimates using those weights would vary from the ground truth estimates when the weights are computed with non-noisy ACS data. With this in mind, we choose to report the results of subgroup estimates for these two survey items. We compute estimates for both items broken out by each race category and again by each education category.

Since we are using the Laplace mechanism to add noise to the ACS cell counts, we compare the weighted estimate obtained from the differentially private release of the ACS data to the weighted estimate obtained from the true ACS data for different values of the privacy loss parameter, ϵ. For each value of ϵ we run 30 experiments. The results of our simulations are summarized in Figures 1-4.

For the largest demographic subgroups (e.g., whites and people with no high school education), the effect of differential privacy on these estimates is negligible. Even with an ϵ of 0.1, the noise added to the ACS cell counts does not affect the post-stratification weights drastically enough to change Clinton's estimated lead among whites (Figure 1) or net support for an assault rifle ban among those without a high school education (Figure 4). This is the behavior we expected here since applying differential privacy is not expected to dramatically affect summary statistics when there is a large sample size.

For slightly smaller subgroups, we do observe some shifts in the CCES estimates for smaller values of epsilon. For instance, using an ϵ of 0.1 to add Laplace noise to the ACS cell counts and then weighting the CCES to this data moves the estimate of Clinton's lead among black Americans by roughly a percentage point (Figure
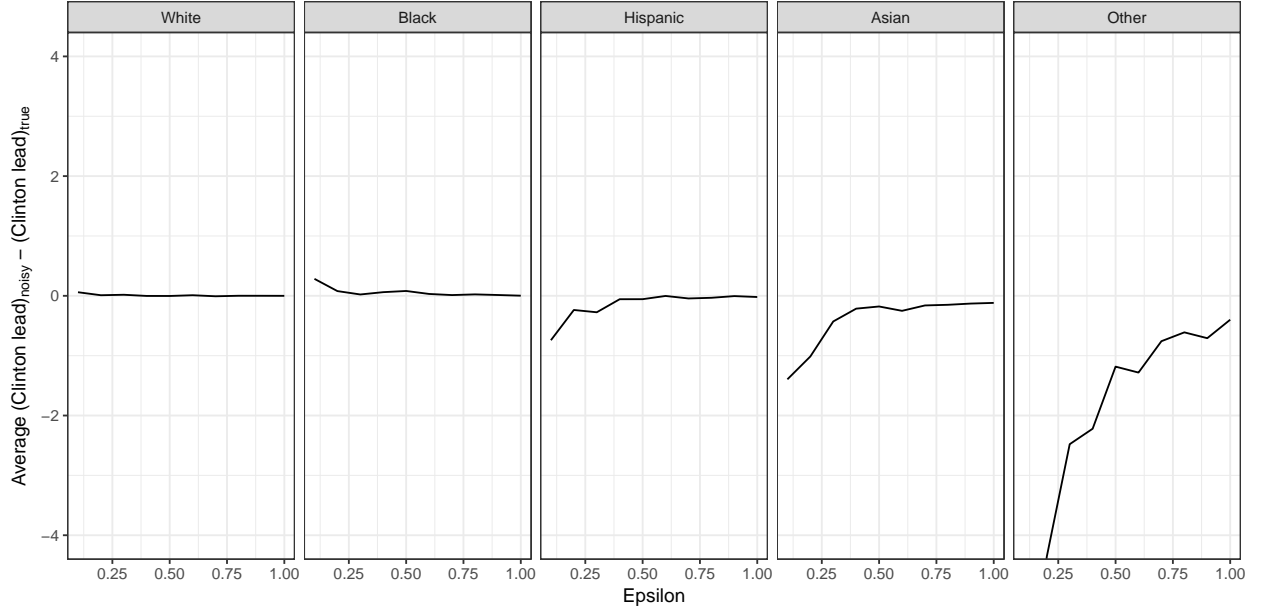
Figure 1: Evaluation of estimates of Clinton's lead by race in the CCES using weights computed from hypothetical private and non-private ACS releases for various values of epsilon.
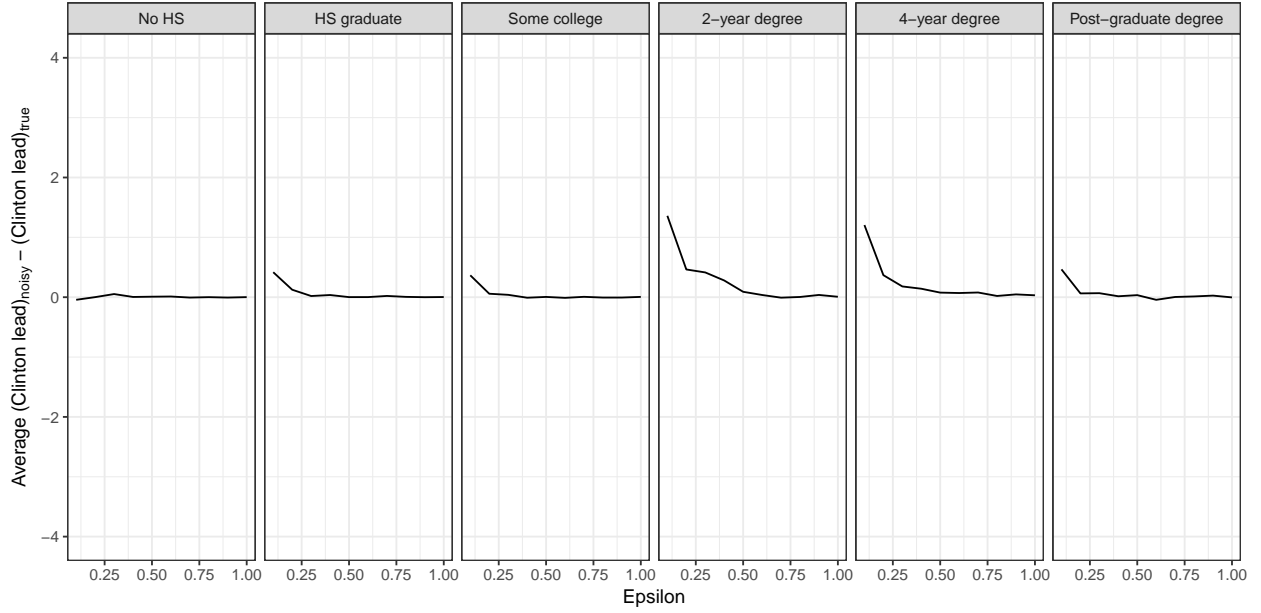

Figure 2: Evaluation of estimates of Clinton's lead by education in the CCES using weights computed from hypothetical private and non-private ACS releases for various values of epsilon.

1). Similarly, with an epsilon value of 0.1, the estimate of net support for an assault rifle ban among those with a 4-year degree (who account for about one-third of the U.S. population) shifts by about one percentage point when the post-stratification weights come from the noisy ACS data (Figure 4). As we increase the value of epsilon and decrease the amount of noise we add to any cell count, the difference between the estimate obtained from weighting the survey to the private and non-private population data goes to zero.

Since these plots measure the average difference in differences, they capture bias that gets introduced into the
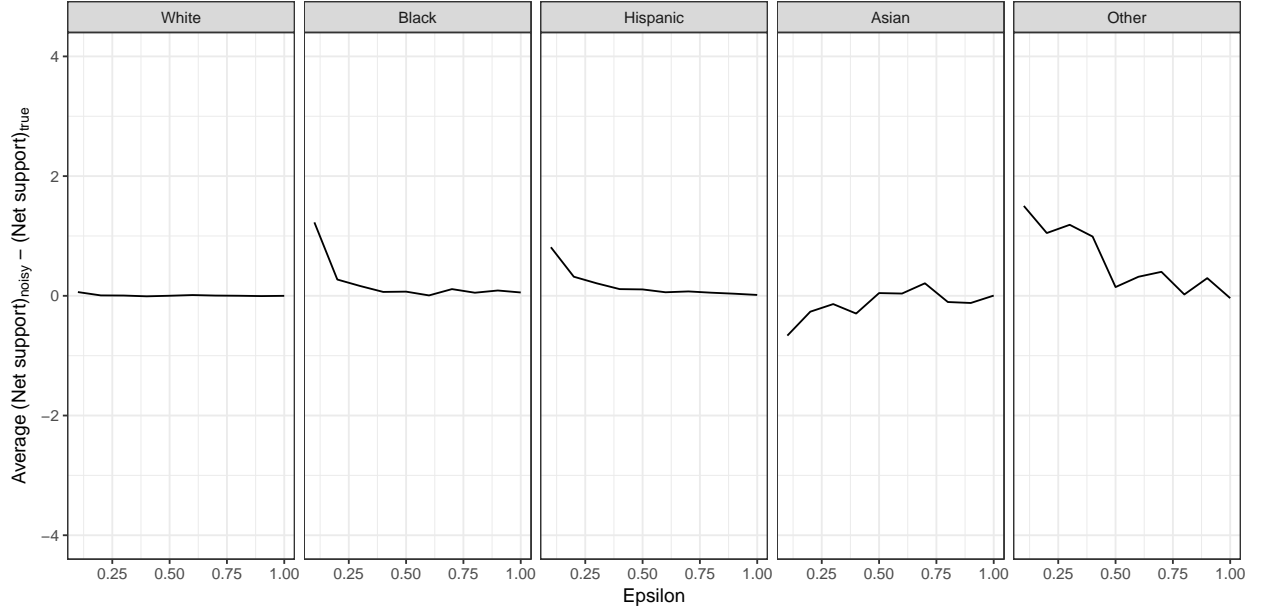
6

Figure 3: Evaluation of net support for assault rifle ban by race in the CCES using weights computed from hypothetical private and non-private ACS releases for various values of epsilon.
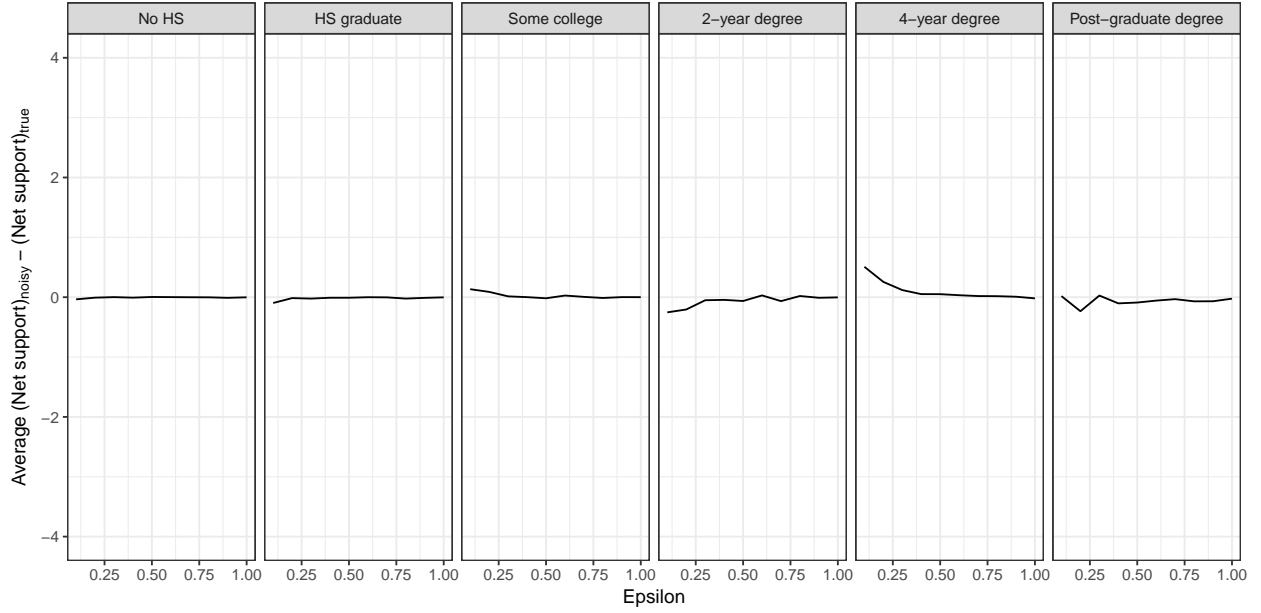

Figure 4: Evaluation of net support for assault rifle ban by education in the CCES using weights computed from hypothetical private and non-private ACS releases for various values of epsilon.

estimates of the survey items (see Appendix A2 for a look at the variance of these estimates). For instance, Figure 1 shows that, for small values of ϵ, Clinton's lead among Asians when the CCES is weighted to the noisy ACS release is smaller than it is when it is weighted to the true ACS. This suggests that releasing a differentially private version of the ACS results in a bias toward Trump in this estimate. Why might this be happening?

We hypothesize that CCES respondents who have combinations of demographic variables that are relatively

7

uncommon in the population have a larger expected upweighting factor when we weight the CCES to a noisy release of the ACS cell counts. When we add noise drawn from the symmetric Laplace distribution, the expected value of the noisy cell counts from small initial cell counts is larger than the initial count because we enforce bin positivity whereas the expected value for larger initial cell counts is equal to the initial cell count. If these respondents who are more likely to get upweighted also have systematically different preferences than respondents whose combinations of demographic variables, then we add bias when we upweight their responses and compute population-level estimates.

For instance, using the sensitivity we derived in *3.2* and an ϵ value of 0.5, we add or subtract a maximum of roughly 20,000 to the ACS cell counts. For a large cell, say one with a population of 100,000, the expected value of the noisy cell count is still 100,000. But for a cell with 10,000, the expected noisy cell count is greater than 10,000 since we enforce bin positivity. This means that, on average, we will upweight the responses of respondents who fall into that cell. If the preferences of those who fall into the cell that has 10,000 are different than the preferences of those who fall into the cell with 100,000, than we bias the estimate toward the preferences of those from the smaller bin.

For a concrete example, let's return to Figure 1. It appears that we add a pro-Clinton bias for black Americans and a pro-Trump bias for Asian Americans. To test our hypothesis, we look at the unweighted estimate of Clinton's lead among blacks who have combinations of demographic variables with fewer than 10,000 people in the ACS versus those who have combinations of demographic variables with greater than 10,000 people in the ACS. We find that blacks who come from rarer subpopulations favor Clinton by about 70 points while blacks who come from more common subpopulations favor Clinton by about 68 points. We repeat this analysis for Asian Americans and find that those who come from rarer subpopulations favor Clinton by about 30 points while those who come from more common subpopulations favor her by 36 points.

This empirical finding is consistent with our hypothesis and appears to hold across the other combinations of demographics and survey items we consider. Thus, it appears that differential privacy may add bias into downstream estimates in this context when subgroups who are rarer in the population have systematically different preferences than more common subgroups. But this bias is predictable.

5. Conclusion

Our analysis demonstrates that with the addition of a reasonable amount of Laplace noise, population-level estimates of presidential candidate preference and support for gun control do not change much for various large demographic subgroups. However, for small values of ϵ the estimates move by as much as one percentage point for some subgroups. On its own, this may not be particularly meaningful, especially since survey researchers are accustomed to dealing with other forms of error, including sampling and measurement error. But this privacy error, even if it is small, is still a new type of error that researchers will need to incorporate into their estimates. Furthermore, in the survey weighting context, there does seem to be a predictable bias when differential privacy affects estimates: the preferences of those who come from smaller subpopulations are upweighted.

We think that it is important to contextualize this analysis by its relative strengths and limitations. The concern we have expressed is that adding noise to population data would affect the relative count of small subgroup populations. But isolating just race or just education in the CCES does not leave us with particularly small sub-populations, which means that we might observe more substantial and meaningful shifts for combined education and race groups, for example. Our analysis, however, is a stylized one due to uncertainty that exists in how the ACS will implement differential privacy and to our desire to simplify the CCES sampling and weighting schemes for illustrative purposes in this paper. We choose not to evaluate these smaller subgroups out of the concern that any effects we discover would be artifacts of our stylized simulation and not of the expected effects of differential privacy on statistical calibration techniques in general. We think any effects observed among these large subgroups are meaningful.

In our simulations we only test $\epsilon \in (0, 1]$. Most existing deployments of differentially privacy have used $\epsilon >> 1$. If the Census follows suit, it seems likely that the effects of differential privacy will be negligible,

as the difference between the noisy weighted estimated and true weighted estimate trends toward 0 as ϵ approaches 1. This can be seen in Figures 1-4.

Additionally, we found that determining the correct scale of noise to add to the ACS cell counts that constitute our population joint distribution was challenging. Throughout the course of this project, we learned that our workaround for the sensitivity calculation is much closer to a reasonable sensitivity that the Census Bureau hopes to attain after updating its ACS processing pipeline than a worst case sensitivity for the current ACS. It is entirely possible that the ACS will effectively add either more or less noise than we do in this analysis. Our findings should not be interpreted as bounding the effects of differential privacy on survey weighting, but as providing one plausible estimate of the effect. Future work should improve upon our fairly naive sensitivity calculation.

One possible approach to reducing the sensitivity would be to clip the ACS "person weights," especially the larger weights. Clipping is a mechanism that is generally implemented in a differentially private mechanism to decrease the global sensitivity and protect outliers in the dataset. It allows for less noise to be added to the privately-released statistic, which should increase the utility of the statistic. However, by using the maximum person weight of 549 in our global sensitivity for the addition of Laplace noise, we are not protecting outliers in the ACS dataset because there is a long right tail in the distribution of person weights, as seen in Appendix A1. A potential solution to this problem is to use the exponential mechanism to release a differentially private value for the 99th percentile, which we would consider to be the maximum person weight, and then clipping to this value. Implementing a differentially private mechanism for releasing the maximum person weight for the purpose of clipping and studying how clipping to different values for "person weight" biases the ACS counts and cell weights for certain demographics is left for future work. Furthermore, exploring the effect of clipping on specific questions asked in the CCES would be useful to examine the introduction of any biases into the post-stratification weighting.

Finally, the PUMS data released by the ACS is only subset of the population for a certain geographic level (in our case, the state). Ideally, ACS would use a differentially private synthetic data generation process to release the PUMS data. We would have liked to simulate this synthetic data generation process, but this would have required having access to the full ACS data, not just the PUMS data. For now, adding Laplace noise to the demographic cell counts generated from the PUMS data is a workaround to study the effects of a differentially private synthetic data release on surveys that rely on the ACS to adjust the findings in a sample. We hope that this stylized study provides a springboard from which to further explore how a differentially private ACS data product will affect downstream statistical calibration methods.

6. References

[1] American Community Survey. 2014. *Chapter 11: Weighting and Estimation.*

[2] Alexander, Davern, and Stevenson. 2010. *Inaccurate Age and Sex Data in the Census PUMS Files: Evidence and Implications.* Federal Reserve Bank of San Francisco.

[3] Dajani, Lauger, Singer, Kifer, Reiter, Machanavajjhala, Garfinkel, Dahl, Graham, Karwa, Kim, Leclerc, Schmutte, Sexton, Vilhuber, Abowd. 2017. *The modernization of statistical disclosure limitation at the U.S. Census Bureau.*

[4] Mercer, Lau, and Kennedy. 2018. "For Weighting Online Opt-In Samples, What Matters Most?" Pew Research Center. Accessed from: https://pewrsr.ch/2HdVXfb.

[5] Shlomo, Krenzke, and Li. 2018. *Comparison of Post-tabular Confidentiality Approaches for Survey Weighted Frequency Tables.*

[6] Wood, Alexandra, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, et al. 2018. *Differential Privacy: A Primer for a Non-Technical Audience.* Vanderbilt Journal of Entertainment & Technology Law 21 (1): 209.

Appendix

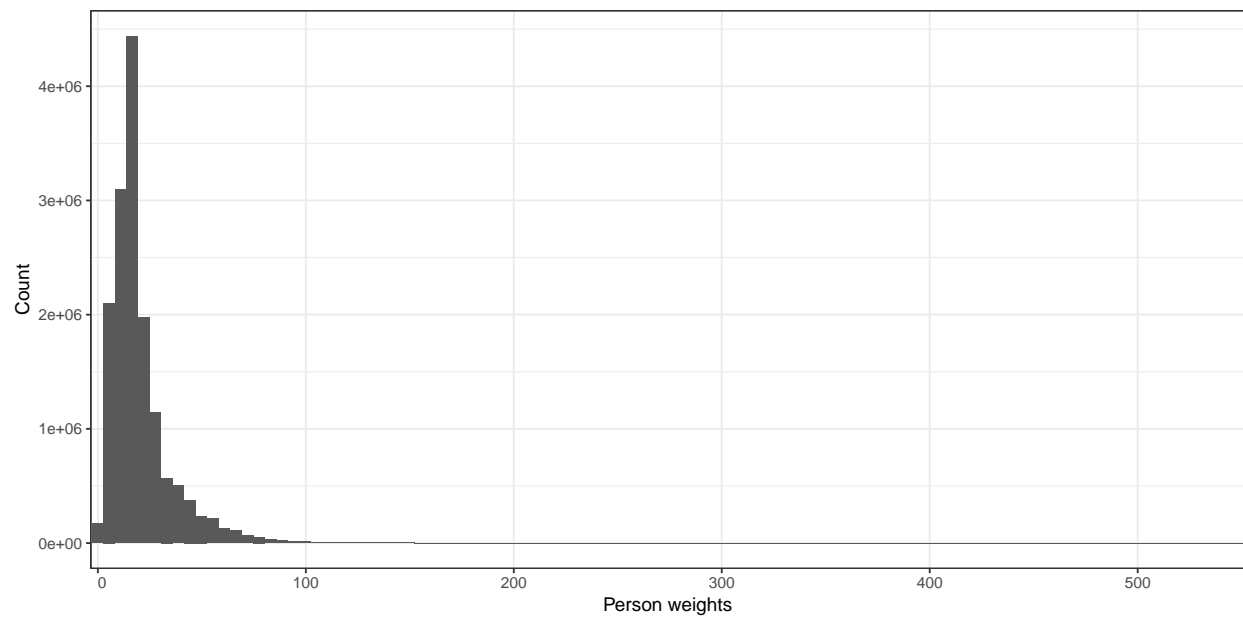## A1. Distribution of person weights in ACS estimates.

Figure 5: Distribution of person weights in ACS estimates.

A2. Root mean squared error between estimates weighted to differentially private and true ACS
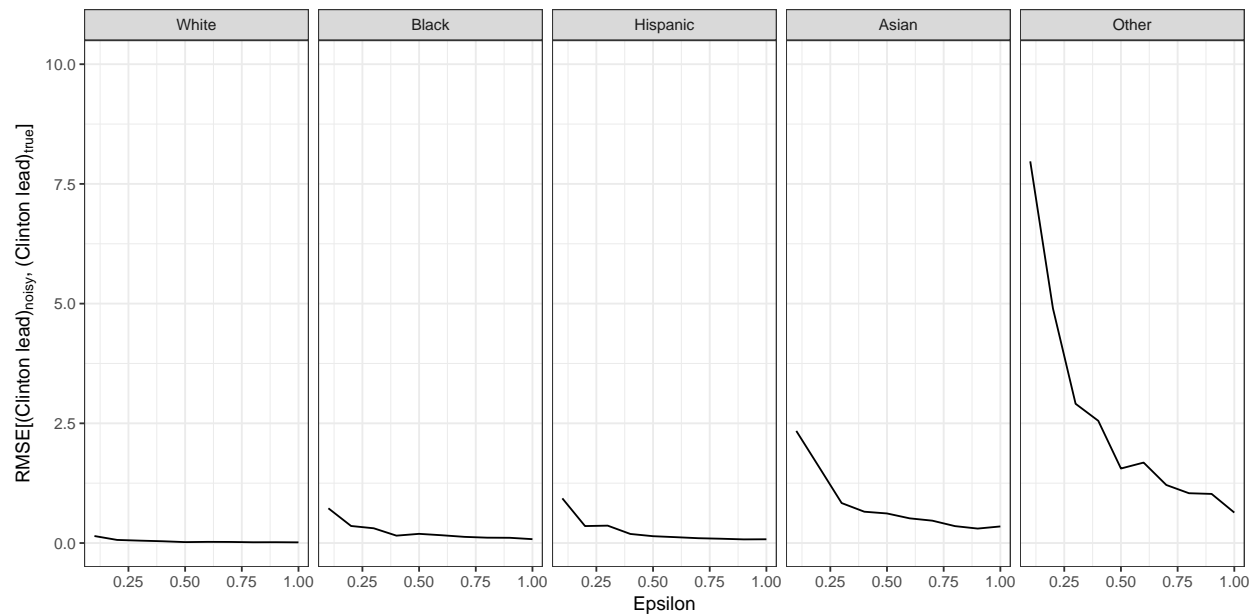

Figure 6: RMSE between estimate of Clinton's lead by race when weighted to differentially private ACS release and estimate when weighted to true ACS.
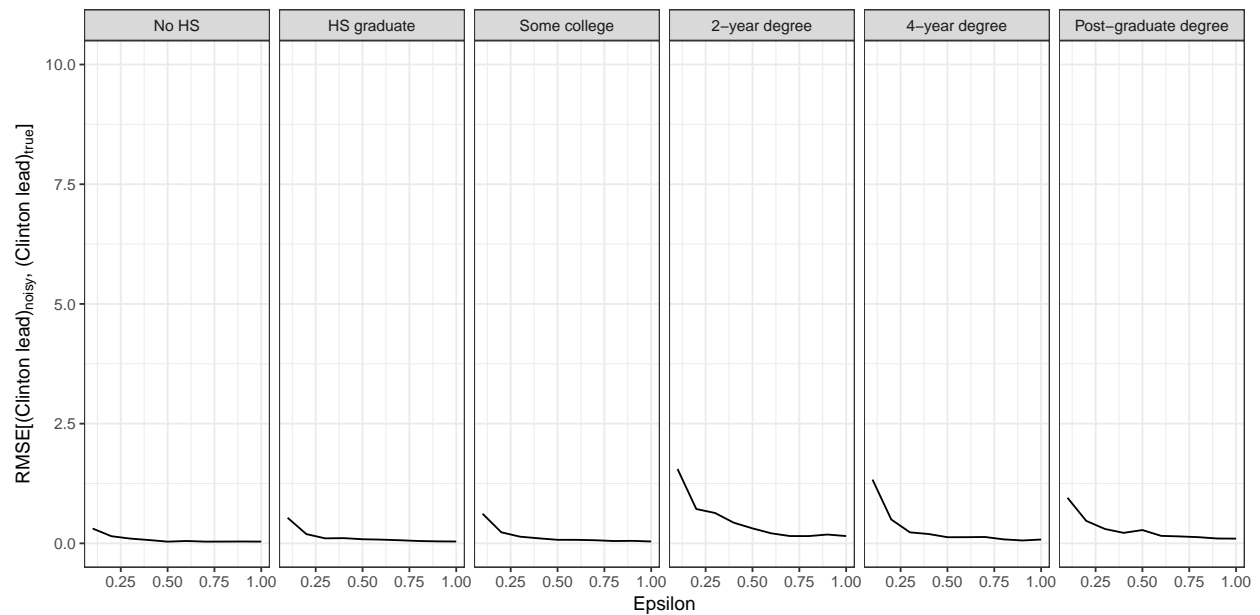

Figure 7: RMSE between estimate of Clinton's lead by education when weighted to differentially private ACS release and estimate when weighted to true ACS.
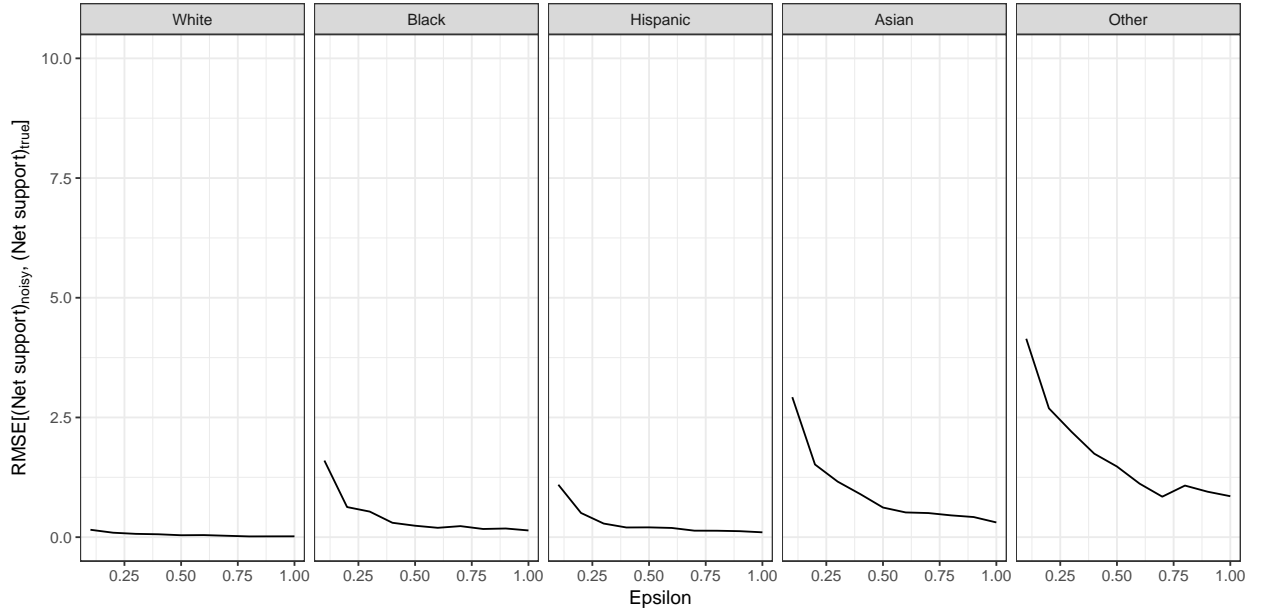
Figure 8: RMSE between estimate of net support for assault rifle ban by race when weighted to differentially private ACS release and estimate when weighted to true ACS.
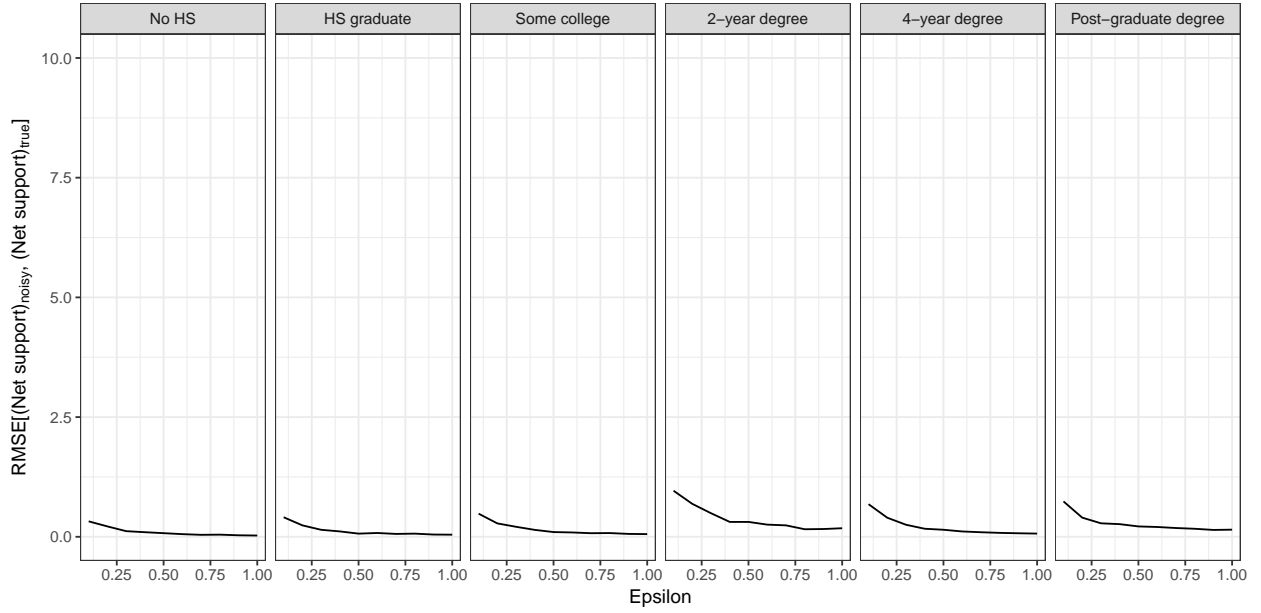

Figure 9: RMSE between estimate of net support for assault rifle ban by education when weighted to differentially private ACS release and estimate when weighted to true ACS.