# Differentially Private Population Quantity Estimates via Survey Weight Regularization (Draft)

Jeremy Seeman
Michigan Institute for Data Science and Institute for Social Research
University of Michigan, Ann Arbor, MI,
48104, USA
jhseeman@umich.edu

Yajuan Si
Institute for Social Research
University of Michigan, Ann Arbor, MI,
48104, USA
yajuan@umich.edu

Jerome P. Reiter
Department of Statistical Science
Duke University, Durham, NC
27708, USA
jreiter@duke.edu

May 5, 2024

**Abstract**

In general, it is challenging to release differentially private (DP) versions of survey-weighted statistics with low error for acceptable privacy loss. This is because weighted statistics from complex sample survey data can be more sensitive to individual survey responses than unweighted statistics. However, when weighted

and unweighted statistics are similar, privacy-preserving noise can dominate any bias corrections for population representation. In this paper, we formalize this three-way trade-off between bias, precision, and privacy; doing so demonstrates the provable limitations of using survey weights as-is in DP analyses. To remedy this, we present a DP method for estimating finite population quantities by first, privately estimating a hyperparameter that determines how much to regularize or shrink survey weights as a function of privacy loss. By adaptively navigating the three-way trade-off of bias variance, and privacy, for each population quantity estimator, we release statistics with sensitivity tailored to the particular relationship between survey weights and response variables. We illustrate the DP finite population estimation using the Panel Study of Income Dynamics, showing that optimal strategies for releasing DP survey-weighted mean income estimates require orders-of-magnitude less DP noise than naively using the original survey weights without modification.

**Keywords:** differential privacy; survey statistics

# 1 Introduction

Privacy protection is crucial for the public release of complex sample survey data. Existing applications of differential privacy (DP) [11, 12] are typically based on censuses or administrative data, failing to account for the data collection process. Survey weights are regularly available to adjust for the unequal probability of individual selection and balance the sample decomposition with the target population. Weighted statistics offer unbiased and consistent estimates for finite population quantities, such as the overall population mean. Our goal is to apply DP to complex sample surveys in releasing survey-weighted statistics.

Preliminary work at the intersection of DP and survey statistics has focused on synthetic data generation including weights [16], survey statistics under classical sampling designs like stratified sampling [19], and privacy interpretations of existing survey sampling methods for their privacy amplification properties [7, 15]. Each of these approaches attempts to utilize as much information as possible about the sampling process. However, weighting schemes can cause practical problems for DP, as weighted statistics have significantly larger sensitivities than their unweighted counterparts, hence requiring substantially more noise to provide the equivalent level of DP protections at the sample level of privacy loss [10]. Moreover, while survey weights may theoretically correct for selection bias, weighted and unweighted estimates may be quite similar if the weights are uncorrelated with a particular survey outcome [6]. Additionally, anomalously large survey weights can significantly increase the variability of survey statistics, prompting approaches to smoothing the estimates or regularizing survey weights [13, 4].

Introducing DP into survey inference suggests that the degree of weight regularization should depend on privacy loss budgets; just as there is a bias-variance-privacy trilemma for mean estimation with independently identically distributed (iid) records [17], similar three-way trade-offs must be made when analyzing survey-weighted quantities. Our work proposes methods for DP survey-weighted estimates where the optimal degree of regularization depends on the confidential data. In this setting, we must consume privacy loss to estimate this optimal degree of regularization *and* the statistics of interest. Doing so allows us to adaptively consume privacy loss budget for fine-tuning uncertainty quantification when constructing interval estimates and assessing their coverage properties.

## 1.1 Contributions

We summarize our contributions here:

1. In Section 3, we analyze the three-way relationship between privacy loss, accuracy, and bias emerging from survey data. To do this, we introduce a regularization parameter $\lambda \in [0, 1]$ that linearly shrinks the survey weights to uniform when $\lambda = 1$. For any survey sample, there exists an "optimal" value $\lambda^*$ which minimizes DP mean-square error (for a fixed privacy loss) that depends on the sample size, response range, possible weighting designs, and the difference between the unweighted and weighted mean estimates. We prove that $\lambda^* > 0$ (for any non-trivial sampling design); similarly, we prove that for any fixed privacy loss, there is a limit to the amount of bias that can be corrected by design-based weight adjustment without requiring DP noise that exceeds said correction.

2. In Section 2.4, we propose a two-step procedure to estimate survey-weighted population means using $\rho$-zero-concentrated Differential Privacy $\rho$-zCDP [8]. First, we use the exponential mechanism to estimate $\lambda^*$; then we use this output to shrink the survey weights and estimate the population mean using the Gaussian mechanism. We also provide different asymptotic and finite-sample approaches to quantifying errors due to sampling, weight shrinkage, and DP noise, allowing users to construct DP confidence intervals for our population mean estimates.

3. In Section 3, we demonstrate our methodology on survey microdata from the Panel Study of Income Dynamics (PSID) [24], an economic survey of families designed to oversample from lower income sub-populations. We show how different response variables require different degrees of survey weight regularization, allowing us to more efficiently tailor DP privacy loss budgets when estimating multiple population means for different response variables. We also empirically validate our uncertainty quantification properties, including accuracy and coverage.

## 1.2 Related Literature

While there's an extensive literature on differentially private statistical analyses (see [23] for a review) and a separate literature on methods for shrinking, trimming, or otherwise regularizing survey weights [14], there is little literature at their intersection. Many DP algorithms rely on the "amplification by sub-sampling" property, wherein applying a DP algorithm on a simple random sample without replacement yields smaller privacy loss than the same algorithm applied to the entire population [3]. However, for survey designs besides simple random sampling, this property may not hold [7] nor would it always improve accuracy [15]. Our work, alternatively, only considers design-based sampling where the weights themselves contain all relevant sampling information and, therefore, must be protected with DP. We do not consider the release of auxiliary data used to construct design-based weights, instead isolating the privacy cost of incorporating survey design exclusively within the weights.

The most direct line of work compared to ours uses methods that jointly generate synthetic data samples containing survey responses and weights [16]. While these methods can produce synthetic data that's interoperable with existing analyses and admits combining-rules-based approaches to synthetic data, our approach differs in a few key ways. First, our work provides finite-sample privacy and accuracy guarantees that do not rely on combining rules which require multiple replicates of the synthetic data. Second, our work provides decision-making guidelines for whether certain kinds of weighting corrections can be sufficiently estimated using DP at a given sample size. Because the underlying synthetic data models depend on more granular (i.e., sensitive) statistics than those based on our estimates, privacy regimes where our methodology fails necessarily implies DP synthetic data methods also fail.

# 2 Methods

## 2.1 Notation and Problem Definition

We consider a response variable and survey weights $\{(y_i, w_i)\}_{i=1}^N$ lying within bounded intervals $[L_Y, U_Y] \times [L_W, U_W]$ from a population of $N$ observations, where we observe the first $i \in [n]$ units and we do not observe $i \in \{n+1, n+2, \ldots, N\}$. For convenience, we define $\Delta_W \triangleq U_W - L_W$, and without loss of generality, we assume $L_Y = 0$ and $1 \leq L_W \leq U_W$. We will use $\boldsymbol{y}$ and $\boldsymbol{w}$ to correspond to the vector of $n$ observed samples and weights, respectively.

Our goal is to estimate the population mean $\theta \triangleq \frac{1}{N}\sum_{i=1}^N y_i$ using the survey-weighted mean

$$\hat{\theta}(\boldsymbol{y}, \boldsymbol{w}) \triangleq \frac{1}{N}\sum_{i=1}^{n} y_i w_i \tag{1}$$

assuming we only have access to the survey responses and the weights. The variability of $\hat{\theta}$ about $\theta$ depends on our sampling mechanism, which we assume is fully characterized by the survey weights (also known as the "design-based" setting in in the survey literature). This allows us to treat the $y_i$s and $w_i$s as fixed constants, making our analysis consistent with DP approaches that treat confidential data entries as constants from a fixed "schema" of possible values.

We focus on the case where the weights correspond to probabilities of inclusion, i.e. $w_i^{-1} \triangleq \pi_i = \mathbb{P}(I_i = 1)$. Doing so yields the classical Horvitz-Thompson variance estimator

$$\widehat{\mathrm{Var}}_{\mathrm{HT}}[\hat{\theta}] \triangleq \frac{1}{N^2}\left(\sum_{i=1}^{n}\frac{1-\pi_i}{\pi_i^2}y_i^2 + \sum_{i=1}^{n}\sum_{j\neq i}\left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\frac{y_i y_j}{\pi_{ij}}\right)\right), \tag{2}$$

where $\pi_{ij} = \mathbb{P}(I_i = 1, I_j = 1)$ is the joint probability of selecting units $i$ and $j$. When the second term in (2) is negative or 0, as can be the case in with-replacement sampling, a conservative approximation uses only the first term in (2), yielding our simplified estimator

$$\widehat{\mathrm{Var}}_{\mathrm{ApproxHT}}[\hat{\theta}] \triangleq \frac{1}{N^2}\sum_{i=1}^{n}\frac{1-\pi_i}{\pi_i^2}y_i^2. \tag{3}$$

The estimator in (3) is also the unbiased variance estimator of the population mean for data collected by Poisson sampling, for which by design

$$\pi_{ij} = \mathbb{P}(I_i = 1, I_j = 1) = \mathbb{P}(I_i = 1)\,\mathbb{P}(I_j = 1) = \pi_i\pi_j. \tag{4}$$

## 2.2 Privacy Background

Next, we introduce our privacy definition. First, we define our adjacent datasets.

**Definition 1** (Adjacency)**.** *We say that two observed samples of size $n$ are adjacent if and only if they differ on the contributions of one observed record, i.e.,*

$$\{(y_i, w_i)\}_{i=1}^n \sim_M \{(y_i', w_i')\}_{i=1}^n \iff \#\{i \in [n] \mid y_i \neq y_i' \text{ or } w_i \neq w_i'\} \tag{5}$$

Note that our analysis assumes that survey weights are fixed properties of individual records that do not change depending on which units appear in the realized sample. This helps align our analysis with standard DP analyses that treat observed confidential data (in this case, survey responses and weights) as constants instead of random variables. We additionally assume that the population and sample sizes, $N$ and $n$, respectively, are public information. While this assumption reflects standard practice for publishing survey metadata, there may be confidentiality concerns if membership in the population under study is privacy-concerning.

Next, we define our DP distance metrics.

**Definition 2** ($\epsilon$-Differential Privacy [11, 12]). *Let $M$ be a randomized algorithm which releases statistics based on $\{(y_i, w_i)\}_{i=1}^n$. We say that algorithm $M$ satisfies $\epsilon$-differential privacy ($\epsilon$-DP) if, for all adjacent observed samples,*

$$d_\infty \left( M(\boldsymbol{y}, \boldsymbol{w}) \mid\mid M(\{\boldsymbol{y}', \boldsymbol{w}'\}) \right) \leq \epsilon, \tag{6}$$

*where above, $d_\infty(\cdot \mid\mid \cdot)$ is the log-max divergence.*

**Definition 3** ($\rho$-zero-concentrated Differential Privacy [8]). *Let $M$ be a randomized algorithm which releases statistics based on $\{(y_i, w_i)\}_{i=1}^n$. We say that algorithm $M$ satisfies $\rho$-zero-concentrated differential privacy ($\rho$-zCDP) if, for all adjacent observed samples and all $\alpha \in (1, \infty)$,*

$$d_\alpha \left( M(\boldsymbol{y}, \boldsymbol{w}) \mid\mid M(\{\boldsymbol{y}', \boldsymbol{w}'\}) \right) \leq \rho\alpha, \tag{7}$$

*where above, $d_\alpha(\cdot \mid\mid \cdot)$ is the $\alpha$-Reyni divergence.*

**Lemma 1** ([8]). *If $M$ satisfies $\epsilon$-DP, it also satisfies $\frac{\epsilon^2}{2}$-zCDP.*

To interpret Equation 7, DP provides numerous *semantic* privacy guarantees about the ability for adversaries to distinguish between two adjacent databases under various definitions of adjacency and data generating processes; see [18] for an example discussion of these semantics. Many of these guarantees rely on independence, i.e., assuming that data comes from an independent (but not necessarily identically) distributed process. Equation 4 ensures that these semantic privacy guarantees hold because we assume that modifying one survey weight does not affect the others, both for our privacy analysis and our data generating assumption analysis.

Next, we briefly introduce some mechanisms which satisfy $\rho$-zCDP, Our method combines two common base algorithms used to satisfy DP: the Gaussian mechanism [8] and the exponential mechanism [20]. We first introduce the Gaussian mechanism.

**Definition 4** (Gaussian Mechanism [8]). *Suppose the statistic*

$$T : \{[L_Y, U_Y] \times [L_W, U_W]\}^n \mapsto \mathbb{R}$$

*has sensitivity defined by*

$$\Delta(T) \triangleq \sup_{(\boldsymbol{y}, \boldsymbol{w}) \sim (\boldsymbol{y}', \boldsymbol{w}')} \left| T(\boldsymbol{y}, \boldsymbol{w}) - T(\boldsymbol{y}, \boldsymbol{w}) \right|. \tag{8}$$

*Then the mechanism $M$ defined as*

$$M(\boldsymbol{y}, \boldsymbol{w}) = T(\boldsymbol{y}, \boldsymbol{w}) + \varepsilon, \qquad \varepsilon \sim N\left(0, \frac{\Delta(T)^2}{2\rho}\right), \tag{9}$$

*satisfies $\rho$-zCDP.*

Next, we introduce the exponential mechanism. This mechanism satisfies $\epsilon$-DP, so we use the conversion in Lemma 1 to modify its form.

**Definition 5** (Exponential Mechanism [20]). *Suppose the goal is to minimize a real-valued loss function $\ell$ over output space $\mathcal{Z}$,*

$$\ell : \mathcal{Z} \times \{[L_Y, U_Y] \times [L_W, U_W]\}^n \times \mapsto [0, \infty).$$

*Similarly, we define a functional analogue of the sensitivity given by*

$$\Delta(\ell) \triangleq \sup_{(\boldsymbol{y},\boldsymbol{w})\sim(\boldsymbol{y}',\boldsymbol{w}'),z\in\mathcal{Z}} |\ell(z;\boldsymbol{y},\boldsymbol{w}) - \ell(z;\boldsymbol{y}',\boldsymbol{w}')|. \tag{10}$$

*Then releasing one sample from the distribution over $\mathcal{Z}$ with density given by*

$$f(z) \propto \exp\left(-\frac{\sqrt{2\rho}}{2\Delta(\ell)}\ell(z;\boldsymbol{y},\boldsymbol{w})\right) \tag{11}$$

*satisfies $\rho$-zCDP.*

## 2.3 Bias-Variance-Privacy Trilemmas for DP Population Estimates

In this section, we propose our DP algorithm for releasing population-level survey weighted estimates and their intervals. Full derivations of results are available in Appendix A.

To start, suppose we wanted to naively implement the Gaussian mechanism. In this setting, the sensitivity of the weighted estimator is given by

$$\Delta(\hat{\theta}) = \sup_{(\boldsymbol{y},\boldsymbol{w})\sim_M(\boldsymbol{y}',\boldsymbol{w}')} \left|\hat{\theta}(\boldsymbol{y},\boldsymbol{w}) - \hat{\theta}(\boldsymbol{y}',\boldsymbol{w}')\right| = \frac{U_W U_Y}{N}. \tag{12}$$

So to satisfy $\rho$-zCDP, we could release

$$\hat{\theta}^{(\rho-\text{zCDP})}(\boldsymbol{y},\boldsymbol{w}) \triangleq \hat{\theta}(\boldsymbol{y},\boldsymbol{w}) + \varepsilon, \qquad \varepsilon \sim N\left(0, \frac{\Delta(\hat{\theta})^2}{2\rho}\right). \tag{13}$$

This naive approach requires Gaussian noise with variance that grows with $U_W^2$, which could be prohibitively expensive if $U_W$ is large (i.e., if some units have a significantly larger survey weight than others). Such issues are especially pronounced for surveys, where typical sample sizes are much smaller than those used for DP evaluations [10].

To motivate an alternate approach, let $\hat{\theta}_0$ be the *unweighted* sample mean, allowing us to suggestively rewrite

$$\hat{\theta} = \hat{\theta}_0 + \text{Sign}(\hat{\theta} - \hat{\theta}_0)|\hat{\theta} - \hat{\theta}_0| \tag{14}$$

The first term in the estimand is the standard, low-sensitivity unweighted mean for which classical DP release mechanisms offer optimal utility guarantees [2]. The second term in the estimand contains two components which we call the *weighting bias sign* and *absolute weighting bias* (AWB), respectively. The AWB's high sensitivity makes DP survey estimation difficult in practice; however, the actual *value* of AWB can be quite close to zero for many response variables. Only when survey response variables and survey weights are highly correlated do we see large AWB values; when the two are less correlated, the AWB can be quite small. Therefore we should consider not only whether it's possible to inflate statistic sensitivities to accommodate survey weighting, but whether such an inflation significantly changes our resulting inferences.

Alternatively, consider a regularization parameter $\lambda \in [0, 1]$ that reduces our estimand's dependence on AWB in the form

$$\hat{\theta}_\lambda \triangleq \hat{\theta}_0 + \lambda \ \text{Sign}(\hat{\theta} - \hat{\theta}_0)|\hat{\theta} - \hat{\theta}_0| \tag{15}$$

We can interpret this as a linear "shrinking" of our design survey weights towards uniform probabilities of selection: $\lambda = 0$ corresponds to the original weights and $\lambda = 1$ corresponds to uniform weights. We define this using the function $G_\lambda$ so that

$$G_\lambda(\boldsymbol{w}) \triangleq (1 - \lambda)\boldsymbol{w} + \frac{\lambda N}{n}\mathbb{1}_n, \qquad \hat{\theta}_\lambda(\boldsymbol{y}, \boldsymbol{w}) = \hat{\theta}(\boldsymbol{y}, G_\lambda(\boldsymbol{w})) \tag{16}$$

By contrast, the sensitivity of $\hat{\theta}_\lambda$ is given by

$$\Delta(\hat{\theta}_\lambda) = \sup_{(\boldsymbol{y},\boldsymbol{w})\sim_M(\boldsymbol{y}',\boldsymbol{w}')} \left|\hat{\theta}_\lambda(\boldsymbol{y}, \boldsymbol{w}) - \hat{\theta}_\lambda(\boldsymbol{y}', \boldsymbol{w}')\right| = \frac{G_\lambda(U_W)U_Y}{N} \tag{17}$$

The reduced sensitivity admits the new $\rho$-zCDP estimator,

$$\hat{\theta}_\lambda^{(\rho-\text{zCDP})}(\boldsymbol{y}, \boldsymbol{w}) \triangleq \hat{\theta}(\boldsymbol{y}, G_\lambda(\boldsymbol{w})) + \varepsilon, \qquad \varepsilon \sim N\left(0, \frac{\Delta(\hat{\theta}_\lambda)^2}{2\rho}\right). \tag{18}$$

This reduction in sensitivity comes at a cost based on the difference between $\hat{\theta}(\boldsymbol{y}, G_\lambda(\boldsymbol{w}))$ and $\hat{\theta}(\boldsymbol{y}, \boldsymbol{w})$. We give this quantity a name, *mechanism bias*, to quantify bias induced by the DP mechanism, defined as

$$B(\lambda) \triangleq \mathbb{E}_\varepsilon[\hat{\theta}_\lambda^{(\rho-\text{zCDP})}(\boldsymbol{y}, \boldsymbol{w})] - \hat{\theta}. \tag{19}$$

This quantity measures the difference between our biased DP estimator's expectation and our unbiased non-DP estimator. For our proposed regularization strategy, the mechanism

bias is linear in $\lambda$, i.e.

$$
\begin{aligned}
B(\lambda) &= \hat{\theta}(\boldsymbol{y}, G_\lambda(\boldsymbol{w})) - \hat{\theta}(\boldsymbol{y}, \boldsymbol{w}) && (20)\\
&= \frac{1}{N}\left[\sum_{i=1}^{n} \left(G_\lambda(w_i) - w_i\right) y_i\right] && (21)\\
&= \frac{1}{N}\left[\sum_{i=1}^{n} \left((1-\lambda)w_i + \frac{\lambda N}{n} - w_i\right) y_i\right] && (22)\\
&= \lambda\left[\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{N}\sum_{i=1}^{n} y_i w_i\right] && (23)\\
&= \lambda(\hat{\theta}_0 - \hat{\theta}) && (24)
\end{aligned}
$$

where $\hat{\theta}_0$ corresponds to the unweighted mean. Therefore, we can consider the mean-square error introduced by DP as taking the form

$$
\begin{aligned}
\ell(\lambda; \boldsymbol{y}, \boldsymbol{w}) &\triangleq \mathbb{E}_\varepsilon\left[(\hat{\theta}_\lambda^{(\rho-\text{zCDP})} - \hat{\theta})^2\right] && (25)\\
&= \mathbb{E}_\varepsilon\left[(\hat{\theta}_\lambda^{(\rho-\text{zCDP})} - \hat{\theta}_\lambda + \hat{\theta}_\lambda - \hat{\theta})^2\right] && (26)\\
&= \mathbb{E}_\varepsilon\left[(\hat{\theta}_\lambda^{(\rho-\text{zCDP})} - \hat{\theta}_\lambda)^2\right] + (\hat{\theta}_\lambda - \hat{\theta})^2 + 2\,\mathbb{E}_\varepsilon\left[\hat{\theta}_\lambda^{(\rho-\text{zCDP})} - \hat{\theta}_\lambda\right]\left(\hat{\theta}_\lambda - \hat{\theta}\right) && (27)\\
&= \frac{\Delta(\hat{\theta}_\lambda)^2}{2\rho} + B(\lambda)^2 && (28)
\end{aligned}
$$

Equation 25 characterizes a three-way "bias-variance-privacy" trilemma for DP survey estimation. As we reduce the mechanism bias of our survey estimates, we require more additive noise to satisfy $\rho$-zCDP; moreover, this effect becomes more extreme as $\rho$ gets smaller.

If we were able to optimally navigate this trade-off for a fixed value of $\rho$, we could try to minimize $\ell$ as a function of $\lambda$. This yields the following Lemma.

**Lemma 2.** *Consider minimizing the loss function in Equation 25. Then...*

1. *The mean square error in Equation 25 is minimized by*

$$
\lambda^* \triangleq \min\left\{1, \frac{\frac{U_W}{\rho}\left(\frac{U_Y}{N}\right)^2\left(U_W - \frac{N}{n}\right)}{\left[\frac{1}{\rho}\left(\frac{U_Y}{N}\right)^2\left(U_W - \frac{N}{n}\right)^2 + 2(\hat{\theta}_0 - \hat{\theta})^2\right]}\right\} \tag{29}
$$

2. $\lambda^* > 0$ *iff* $U_W > N/n$.

3. $\lambda^* < 1$ *iff*

$$
\left|\hat{\theta}_0 - \hat{\theta}\right| > \sqrt{\frac{U_Y^2}{2\rho N n}\left(U_W - \frac{N}{n}\right)}, \tag{30}
$$

*or, equivalently,*

$$\rho > \frac{U_Y^2}{2(\hat{\theta}_0 - \hat{\theta})^2 N n} \left( U_W - \frac{N}{n} \right). \tag{31}$$

Lemma 2 has an interesting interpretation. First, if the weighting scheme is non-trivial (i.e., if $U_W > N/n$), then is it *never* optimal to use survey weights as-is without some degree of regularization (i.e., $\lambda^* > 0$). Second, if the effect of the mechanism bias introduced by shrinking survey weights is not sufficiently large, or if $\rho$ is sufficiently small relative to $U_Y^2$, then the optimal DP strategy to minimize $\ell(\lambda \cdot; \boldsymbol{y}, \boldsymbol{w})$ is to *ignore* the survey weights entirely (i.e., $\lambda^* = 1$).

To visualize this effect, suppose we are interested in estimating a binary proportion where $y_i \in \{0, 1\}$ and $\Delta_Y = 1$ from a population of $N = 10^8$. We consider varying the sample size $n$, the privacy loss budget for estimating $\rho$, and the weight ratio $U_W n/N$, i.e. the ratio of the maximum survey weight to the uniformly weighted equivalent. In each case, we calculate the value for the AWB, $\left|\hat{\theta}_0 - \hat{\theta}\right|$, where equality is achieved in Equation 30. This value represents the minimum difference in the population proportion with or without using survey weights necessary to consider accounting for the weighting process in a DP estimate. We plot the values in Figure 1. We see that, as expected, DP estimators for the population mean can better incorporate weighting information as the sample size increases, the weight ratio decreases, and as $\rho$ increases. However, when these parameters trend in the opposite direction, it becomes harder to justify incorporating survey weights into the analysis. For example, in the extreme case where the weight ratio is $10^4$ (meaning one respondent can hypothetically contribute up to $10^4$ more to a weighted mean) and $\rho = 1$, we require at least a 10% difference between the weighted and unweighted statistics for a sample of $n = 10^3$ respondents to justify incorporating survey weights. Results like these can help determine the kinds of survey-weighting corrections that are feasible or infeasible to consider with DP.

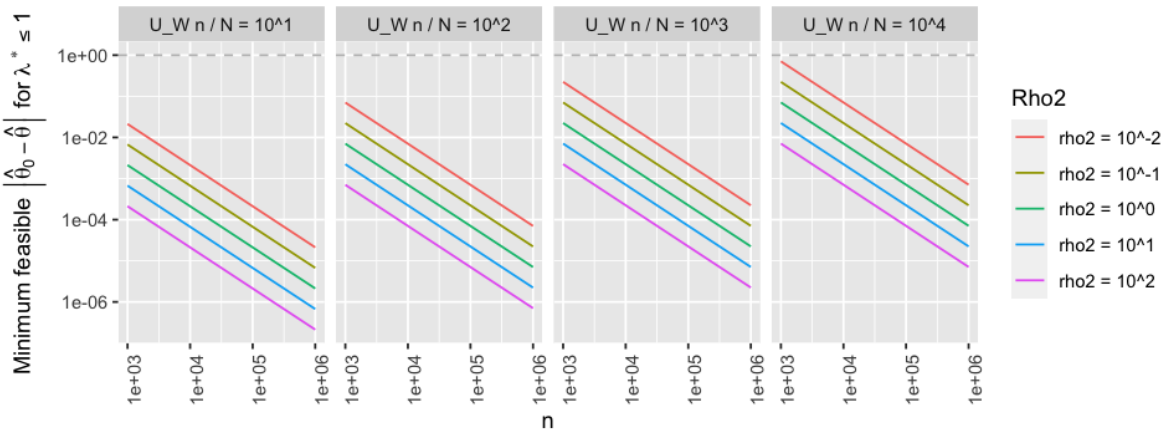



Figure 1: Minimum feasible values for $\left|\hat{\theta}_0 - \hat{\theta}\right|$ as a function of sample size $n$, weight ratio $U_W n/N$, and privacy loss $\rho$, for a population proportion of size $N = 10^8$.

## 2.4 Proposal: DP Survey Weight Regularization

The optimal degree of regularization $\lambda^*$ depends on the confidential data through the AWB, $|\hat{\theta}_0 - \hat{\theta}|$. Therefore, we propose the following two-step approach to estimating $\hat{\theta}$ using $(\rho_1 + \rho_2)$-zCDP, written out in Algorithm 1.

1. Estimate $\lambda^*$ while satisfying $\rho_1$-zCDP by using the exponential mechanism by sampling $\hat{\lambda}^{\rho_1\text{-zCDP}}$ from the density

$$f(\lambda) \propto \mathbb{1}_{\{\lambda \in [0,1]\}} \exp\left(-\frac{\sqrt{2\rho_1}}{2\Delta(\ell)}\ell(\lambda; \boldsymbol{y}, \boldsymbol{w})\right), \tag{32}$$

   where we show $\Delta(\ell) = (\Delta(\hat{\theta}) - \Delta(\hat{\theta}_0))^2$.

2. Using $\hat{\lambda}^{\rho_1\text{-zCDP}}$, sample $\hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}}$ according to the Gaussian mechanism using weights shrunk with estimated optimal lambda value $\hat{\lambda}^{\rho_1\text{-zCDP}}$.

By using $\hat{\lambda}^{\rho_1\text{-zCDP}}$ as a noisy proxy for $\lambda^*$, we still maintain a high probability of reducing the noise needed to satisfy zCDP when adding Gaussian noise to the weighted mean estimate.

**Theorem 1.** *Algorithm 1 satisfies $(\rho_1 + \rho_2)$-zCDP.*

---

**Algorithm 1** DP regularized survey-weighted population estimate

---

**Require:** $\rho_1, \rho_2 \in (0, \infty)$, $\{(y_i, w_i)\}_{i=1}^n \in \{[L_Y, U_Y] \times [L_W, U_W]\}^n$, $N \in \mathbb{N}$.

Sample $\hat{\lambda}^{\rho_1\text{-zCDP}}$ from the density $f(\lambda)$ where

$$f(\lambda) \propto \mathbb{1}_{\{\lambda \in [0,1]\}} \exp\left(-\frac{\sqrt{2\rho_1}}{2\Delta(\ell)}\left(\frac{\Delta(\hat{\theta}_\lambda)^2}{2\rho_2} + B(\lambda)^2\right)\right) \tag{33}$$

Sample $\hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}}$ where

$$\hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}} \sim N\left(\hat{\theta}_{\hat{\lambda}^{\rho_1\text{-zCDP}}}(\boldsymbol{y}, \boldsymbol{w}), \frac{1}{2\rho_2}\left(\frac{1}{N}\left[G_{\hat{\lambda}^{\rho_1\text{-zCDP}}}(U_W)U_Y\right]\right)^2\right) \tag{34}$$

**return** $\begin{pmatrix} \hat{\lambda}^{\rho_1\text{-zCDP}} \\ \hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}} \end{pmatrix}$

---

Next, we discuss errors due to DP using Algorithm 1 by quantifying the concentration around $\hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}}$ about $\hat{\theta}$. The Gaussian mechanism noise error is trivial to quantify. For the mechanism bias, we can lower bound an estimate of $\hat{\lambda}^*$ using $\hat{\lambda}^{\rho_1\text{-zCDP}}$. Since $\lambda^*$ decreases as AWB increases, lower bounding $\hat{\lambda}^*$ with high probability allows us to upper bound AWB with high probability. This yields the following result.

**Theorem 2.** *Let* $\begin{pmatrix} \hat{\lambda}^{\rho_1\text{-}zCDP} \\ \hat{\theta}^{\rho_2\text{-}zCDP}_{\hat{\lambda}^{\rho_1\text{-}zCDP}} \end{pmatrix}$ *be the output of Algorithm 1. Then we have*

$$\mathbb{P}\left(|\hat{\theta}^{\rho_2\text{-}zCDP}_{\hat{\lambda}^{\rho_1\text{-}zCDP}} - \hat{\theta}| \leq C^*\right) \leq 1-\alpha \tag{35}$$

*where*

$$C^* = \frac{B^-\left(\hat{\lambda}^{\rho_1-\text{zCDP}} + z_{\alpha/4}\sqrt{\sup_{\boldsymbol{y},\boldsymbol{w}}\left[\sigma^{2*}(\boldsymbol{y},\boldsymbol{w})\right]}\right)}{\hat{\lambda}^{\rho_1-\text{zCDP}}} + z_{\alpha/4}\frac{G_{\hat{\lambda}^{\rho_1\text{-}zCDP}}(U_W)U_Y}{N\sqrt{2\rho_2}} \tag{36}$$

*and*

$$B^-(\lambda) \triangleq \sqrt{\frac{1}{2}\left[\frac{\frac{U_W}{\rho_2}\left(\frac{U_Y}{N}\right)^2\left(U_W - \frac{N}{n}\right)}{\lambda} - \frac{1}{\rho_2}\left(\frac{U_Y}{N}\right)^2\left(U_W - \frac{N}{n}\right)^2\right]} \tag{37}$$

*Proof.* First, we quantify the magnitude of the mechanism bias. In the absence of DP, when $\lambda^* \in (0,1)$, we have a bijection between $\lambda^*$ and $|\hat{\theta} - \theta|$ of the form:

$$|\hat{\theta}_0 - \hat{\theta}| = \sqrt{\frac{1}{2}\left[\frac{\frac{U_W}{\rho_2}\left(\frac{U_Y}{N}\right)^2\left(U_W - \frac{N}{n}\right)}{\lambda^*} - \frac{1}{\rho_2}\left(\frac{U_Y}{N}\right)^2\left(U_W - \frac{N}{n}\right)^2\right]} \triangleq B^-(\lambda^*). \tag{38}$$

Using the functional form of $f(\lambda)$, $\hat{\lambda}^{\rho_1\text{-zCDP}}$ follows a truncated normal on $[0,1]$ with variance parameter

$$\sigma^{2*}(\boldsymbol{y},\boldsymbol{w}) \triangleq \left(\frac{\sqrt{2\rho_1}}{\Delta(\ell)}\left[\frac{1}{2\rho_2}\left(\frac{\Delta_Y}{N}\right)^2\left(\frac{N}{n} - U_W\right)^2 + (\hat{\theta}_0 - \hat{\theta})^2\right]\right)^{-1}. \tag{39}$$

Independent of any confidential data, we have

$$\sup_{\boldsymbol{y},\boldsymbol{w}}\left[\sigma^{2*}(\boldsymbol{y},\boldsymbol{w})\right] \triangleq \left(\frac{\sqrt{2\rho_1}}{\Delta(\ell)}\left[\frac{1}{2\rho_2}\left(\frac{\Delta_Y}{N}\right)^2\left(\frac{N}{n} - U_W\right)^2\right]\right)^{-1} \tag{40}$$

As $\lambda^*$ increases, $|\hat{\theta}_0 - \hat{\theta}|$ decreases. Therefore if we lower bound $\lambda^*$ using $\hat{\lambda}^{\rho_1-\text{zCDP}}$ with high probability, we can upper bound $|\hat{\theta}_{\hat{\lambda}^{\rho_1-\text{zCDP}}} - \hat{\theta}|$ with high probability, i.e.

$$\mathbb{P}\left(\lambda^* \geq \hat{\lambda}^{\rho_1-\text{zCDP}} + z_{\alpha/2}\sqrt{\sup_{\boldsymbol{y},\boldsymbol{w}}\left[\sigma^{2*}(\boldsymbol{y},\boldsymbol{w})\right]}\right) \geq 1-\alpha \tag{41}$$

which implies

$$\mathbb{P}\left(|\hat{\theta}_0 - \hat{\theta}| \leq B^-\left(\hat{\lambda}^{\rho_1-\text{zCDP}} + z_{\alpha/2}\sqrt{\sup_{\boldsymbol{y},\boldsymbol{w}}\left[\sigma^{2*}(\boldsymbol{y},\boldsymbol{w})\right]}\right)\right) \geq 1-\alpha, \tag{42}$$

and therefore,

$$\mathbb{P}\left(|\hat{\theta}_{\hat{\lambda}^{\rho_1-\text{zCDP}}} - \hat{\theta}| \leq \frac{B^-\left(\hat{\lambda}^{\rho_1-\text{zCDP}} + z_{\alpha/2}\sqrt{\sup_{\boldsymbol{y},\boldsymbol{w}}\left[\sigma^{2*}(\boldsymbol{y},\boldsymbol{w})\right]}\right)}{\hat{\lambda}^{\rho_1-\text{zCDP}}}\right) \geq 1-\alpha. \tag{43}$$

□

We make a few comments about Theorem 2. First, note that we used a union bound to establish our concentration around the two parameters. Depending on prior beliefs about the relative magnitude of errors due to mechanism bias or Gaussian noise addition, one could place more weight on either component in the concentration inequality. Next, note that we can use our estimated regularization parameter as a noisy "plug-in" proxy for estimating AWB with $\rho_1$-zCDP (with the caveat that this plug-in estimator is high-sensitivity, like our original estimand). To do this, suppose $\hat{\theta}_0 > \hat{\theta}$; then

$$|\hat{\theta}_0 - \hat{\theta}| \approx B^-(\hat{\lambda}^{\rho_1-\text{zCDP}}) \implies \mathbb{E}\left[\hat{\theta}^{\rho_2-\text{zCDP}}_{\hat{\lambda}^{\rho_1-\text{zCDP}}} + \hat{\lambda}^{\rho_1-\text{zCDP}} B^-(\hat{\lambda}^{\rho_1-\text{zCDP}})\right] \approx \hat{\theta} \tag{44}$$

Finally, Theorem 2 does not directly considered the weighting bias sign, $\text{Sign}(\hat{\theta} - \hat{\theta}_0)$, as only the AWB shows up in the proof. If this sign is unknown a priori, one can modestly estimate it using an instantiation of the exponential mechanism. However, for many surveys in practice, the sign of the bias can safely be assumed to be public information. For example, our later data analysis considers cases where it's known a priori that our survey design oversamples families from lower incomes, allowing us to reasonably treat the bias direction as public information.

## 2.5 Sampling Error Accounting

In this subsection, we investigate different approaches to simultaneously quantifying errors due to survey sampling and errors due to DP. These approaches differ in their finite-sample and asymptotic coverage guarantees. We discuss their theoretical differences here, which are further investigated empirically in the next section.

First, we consider a global upper bound on the the distance between $\hat{\theta}$ and $\theta$ using concentration inequalities. Following [9], if there exists $k_y, k_w < \infty$ and integers $p, q$ such that $pq \geq p + 2q$ and

$$\left[\frac{1}{N}\sum_{i=1}^{N}|y_i|^p\right]^{1/p} \leq k_y, \qquad \left[\frac{1}{N}\sum_{i=1}^{N}(w_i\overline{\pi})^q\right]^{1/q} \leq k_w \tag{45}$$

where $\overline{\pi}$ is the average probability of inclusion. Then setting $r \triangleq 1 - 1/p - 1/q$, we have

$$\mathrm{Var}[\hat{\theta}] \leq U_V \triangleq \frac{k_y^2 k_w}{\overline{\pi} N} + k_y^2 k_w \left[ \frac{1}{N^2} \sum_{i \neq j} \left| \frac{\mathbb{P}(I_i = 1 | I_j = 1) - \mathbb{P}(I_i = 1)}{\overline{\pi}} \right|^{1/r} \right]^r. \tag{46}$$

Note that under the same Poisson sampling setting assumption, the second term in Equation 46 reduces to 0 as $I_i$ and $I_j$ are independent. For this, we have global moment bounds of the form

$$\left[ \frac{1}{N} \sum_{i=1}^{N} |y_i|^p \right]^{1/p} \leq U_Y \quad \forall p \in \mathbb{N}, \qquad \left[ \frac{1}{N} \sum_{i=1}^{N} (w_i \overline{\pi})^q \right]^{1/q} \leq \frac{U_W}{L_W} \quad \forall q \in \mathbb{N}. \tag{47}$$

This approach has pros and cons. On the one hand, the formulation gives exact accuracy guarantees that depend exclusively on public quantities, therefore requiring no additional privacy loss expenditure. However, in practice, the constants may be prohibitively large to admit reasonable interval estimates for the population mean. For the small sample sizes in our data analysis, we are in the latter regime.

Alternatively, we consider the asymptotic analysis of our estimator. Under the classical, mild asymptotic conditions in [5], as sample size increases relative to population size, we have $\hat{\theta} \to_D N(\theta, \mathrm{Var}_{\mathrm{HT}}(\theta))$. This allows to construct asymptotically consistent confidence intervals using standard plug-in estimators for the mean and variance.

Because $\widehat{\mathrm{Var}}_{\mathrm{ApproxHT}}[\hat{\theta}]$ depends on the confidential data, we must estimate it with zCDP. By spending an additional $\rho_3$ of privacy loss, we can estimate $\hat{V}_{\rho_3\text{-zCDP}}$ and construct asymptotically consistent confidence intervals using Algorithm 2. Note that we do not want to underestimate the sampling variance of our estimator by using shrunk survey weights. Therefore we do not use the regularization parameter here. Additionally, we introduce a new parameter $\alpha_v \in (0, 1)$ to calculate a $(1 - \alpha_v)\%$ upper bound on the confidential $\widehat{\mathrm{Var}}_{\mathrm{ApproxHT}}[\hat{\theta}]$. This allows us to account for uncertainty in $\hat{V}_{\rho_3\text{-zCDP}}$ when constructing our intervals.

**Algorithm 2** DP regularized survey-weighted population confidence interval

**Require:** $\rho_1, \rho_2, \rho_3 \in (0, \infty)$, $\{(y_i, w_i)\}_{i=1}^n \in \{[L_Y, U_Y] \times [L_W, U_W]\}^n$, $N \in \mathbb{N}$, $\alpha \in (0, 1)$, $\alpha_v \in (0, 1)$.

Sample $\hat{\lambda}^{\rho_1\text{-zCDP}}$ and $\hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}}$ according to Algorithm 1.

Sample

$$\hat{V}_{\rho_3\text{-zCDP}} \sim N\left( \widehat{\mathrm{Var}}_{\mathrm{ApproxHT}}[\hat{\theta}], \ \frac{\Delta(\hat{V})^2}{2\rho_3} \right), \qquad \Delta(\hat{V}) = \Delta(\hat{\theta})^2 \tag{48}$$

**return**

$$\hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}} \pm z_{\alpha/2} \sqrt{\frac{\Delta(\hat{\lambda}^{\rho_1\text{-zCDP}})^2}{2\rho_2} + \hat{V}_{\rho_3\text{-zCDP}} + z_{\alpha_v/2} \sqrt{\frac{\Delta(\hat{V})^2}{2\rho_3}}} \tag{49}$$

**Theorem 3.** *Algorithm 2 satisfies* $(\rho_1 + \rho_2 + \rho_3)$*-zCDP.*

**Theorem 4.** *Let* $\hat{\theta}^*$ *be an arbitrary new population mean estimate, drawn from the same sampling scheme. Under the regularity conditions in [5] as* $N, n \to \infty$

$$\mathbb{P}\left(\hat{\theta}^* \in \left[\hat{\theta}^{\rho_2\text{-}zCDP}_{\hat{\lambda}^{\rho_1\text{-}zCDP}} \pm z_{\alpha/2}\sqrt{\frac{\Delta(\hat{\lambda}^{\rho_1\text{-}zCDP})^2}{2\rho_2} + \hat{V}_{\rho_3\text{-}zCDP} + z_{\alpha_v/2}\sqrt{\frac{\Delta(\hat{V})^2}{2\rho_3}}}\right]\right) \to_P 1-\alpha \quad (50)$$

*Proof.* As $N, n$ to $\infty$ under the conditions in [5], the mechanism bias converges to 0, i.e.

$$\hat{\lambda}^{\rho_1\text{-zCDP}} \to_P 0 \qquad \Longrightarrow \qquad \hat{\theta}_{\lambda_{\rho_1\text{-zCDP}}} \to_P \hat{\theta} \quad (51)$$

Similarly, the sensitivities of $\hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}}$ and $\hat{V}_{\rho_3}$ decrease to 0 as $N, n \to \infty$ under the conditions in [5]. Combining all the above, we have that the interval in Equation 49 converges to the classical non-DP normal approximation interval, yielding the result. □

Note that the asymptotic consistency of our DP confidence interval does not require adjustments for uncertainty in $\hat{\lambda}^{\rho_1\text{-zCDP}}$. Instead, we rely on the estimation of $\hat{\lambda}^{\rho_1\text{-zCDP}}$ to determine the degree to which AWB might affect our inferences. Alternatively, we could use the plug-in bias estimate from Equation 44 in place of our mean estimate to offer a partial, noisy bias correction.

# 3 Data Analysis

## 3.1 Data: Panel Study of Income Dynamics

To demonstrate the methodology above, we apply it to the the Panel Study of Income Dynamics (PSID), a longitudinal survey containing family-level statistics on income sources and other sociodemographic information [24]. We use family-level data published from 2019, consisting of $n = 9420$ families from a population of $N \approx 1.29 \times 10^8$. For the purposes of this evaluation, we treat the provided survey weights as design-based (although the full methodology contains model-based adjustments; see [24] for details). Under this weighting scheme, $U_W = 6 \times 10^4$. Our goal is to estimate population mean quantities from the variables of interest in Table 1 using $\rho$-zCDP. Additionally, we include one synthesized random variable, `bern`, which contains iid Bernoulli draws to simulate a random survey response that's independent of the survey weights by construction.

PSID facilitates research on employment, income, wealth, health, family and child development, and other sociodemographic and economic topics. In order to maintain national representativeness, lower-income families are intentionally oversampled in the survey. We plot the bivariate relationship between survey weights and `inc3` in Figure

| Variable | Description | $U_Y$ | $\hat{\theta}_0 - \hat{\theta}$ |
|---|---|---|---|
| `inc3` | Cube-root family income | 150 | -.67 |
| `pov` | 1 if family income below poverty line, else 0 | 1 | .022 |
| `nf` | Number of family members | 20 | .27 |
| `bernoulli` | iid Bernoulli(.5) r.v. | 1 | .004 |

Table 1: Selected PSID variables

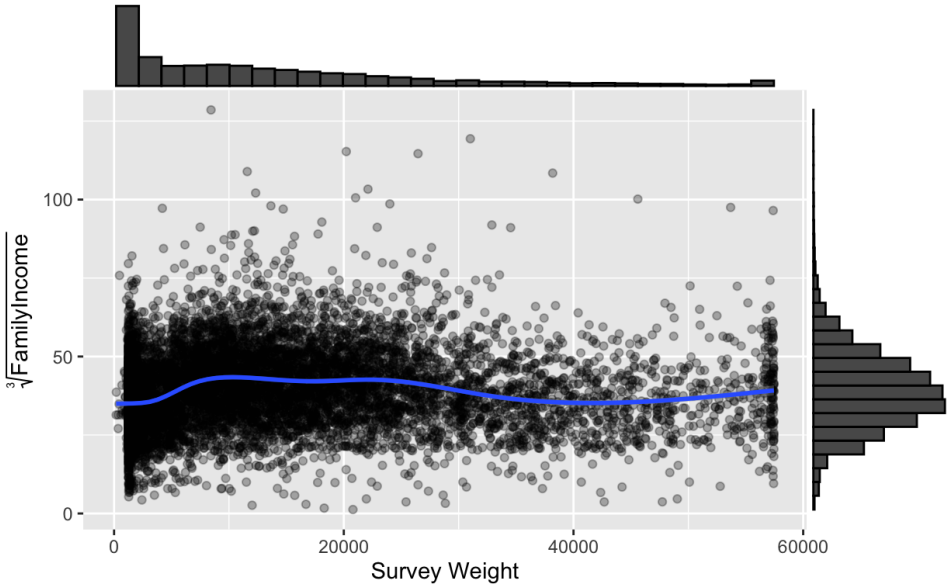



Figure 2: Bivariate scatter plot of survey weights (x-axis) and `inc3` (y-axis), with univariate histograms on the margins and a spline estimate of the central tendency in blue.

2. The final composite survey weights are weakly correlated with family income (Spearman's rank correlation of approximately .14). Similarly, we plot the distribution of survey weights for families below and above the poverty line in Figure 3, where we see a visible difference in distributions. These are reflected in the biases from Table 1; by ignoring survey weights, we would underestimate the national average family income and overestimate the national poverty rate, as expected.

## 3.2 Theoretical Analysis of Privacy-Utility-Bias Trade-Offs

Next, we use the results from Lemma 2 to show how the more privacy loss we are willing to spend, the more fine-grained the AWB corrections can become. Figure 4 is similar to 1 with realized values from the two income-related PSID response variables. For demonstration purposes, we theoretically vary the sample size $n$ (represented by the different colored lines). The $x$-axis represents the privacy loss budget spent on estimating the population mean, and the $y$-axis refers to the smallest possible weighting bias for which $\lambda^* < 1$, i.e. for which we still benefit from accounting for survey weighting in DP inference. As expected, smaller AWB values can be accommodated with larger sample

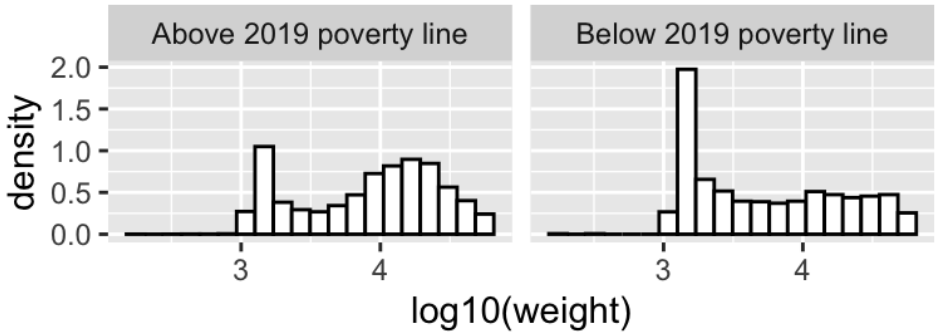



Figure 3: Histograms of survey weights for respondents above and below the 2019 poverty lines (left and right, respectively).

sizes and privacy loss budgets. The horizontal dashed lines refer to the realized AWBs for the two variables of interest (in a pure DP analyses, these would be confidential). Values above these lines refer to realized biases that would admit non-trivial bias corrections at the allowed privacy loss budget level *a priori*.

Next, we visualize the privacy-bias-variance trade-off for the population mean estimates of our different variables. Figure 5 shows how the theoretical bias-variance-privacy trade-off manifests for estimating the survey-weighted average cube-root income (`inc3`) and proportion of families below the 2019 poverty line (`pov`). For comparison purposes, we also include two hypothetical responses: simulated iid Bernoulli(.5) responses (`bernoulli`) and a copy of the survey weights themselves (`wgt`), representing minimal and maximal correlation between survey weights and responses. We plot the noise-to-signal ratio as the theoretical MSE over the weighted mean estimate on the $y$-axis, with the regularization parameter $\lambda$ on the $x$-axis. We see that as the magnitude of the bias decreases (moving from top left subfigure to bottom right subfigure), the optimal MSE is achieved at larger values of $\lambda_{\text{opt}}$ for the same privacy loss budget $\rho_2$. Moreover, as $\rho_2$ decreases, $\lambda_{\text{opt}}$ increases for each response variable under consideration. We see that for reasonably small choices of $\rho_2$, we tend to reject small $\lambda$ to optimize the bias-variance trade-off at each fixed $\rho_2$ value.

Figure 6 shows the sampling distribution of $\hat{\lambda}$ for estimating cube-root income at different values of $\rho_1$ and $\rho_2$. By construction, this statistic is the most sensitive by depending on the gap between the weighted and unweighted population means. Therefore we do not sample $\lambda$ particularly close to the optimal $\hat{\lambda}$ without a large $\rho_1$. However, even for small values of $\rho_1$, we can reasonably avoid sampling small values of $\lambda$ with high probability, which allows us to avoid the worst of the sensitivity inflation in the next stage.

Figure 7 shows the sampling distribution of the plug-in bias estimate $B^{-}(\hat{\lambda}^{\rho_1\text{-zCDP}})$ at different levels of $\rho_1$. As expected, when $\rho_1$ is small, we are more likely sample larger $\hat{\lambda}^{\rho_1\text{-zCDP}}$ values which implies we underestimate the magnitude of the bias. As expected,

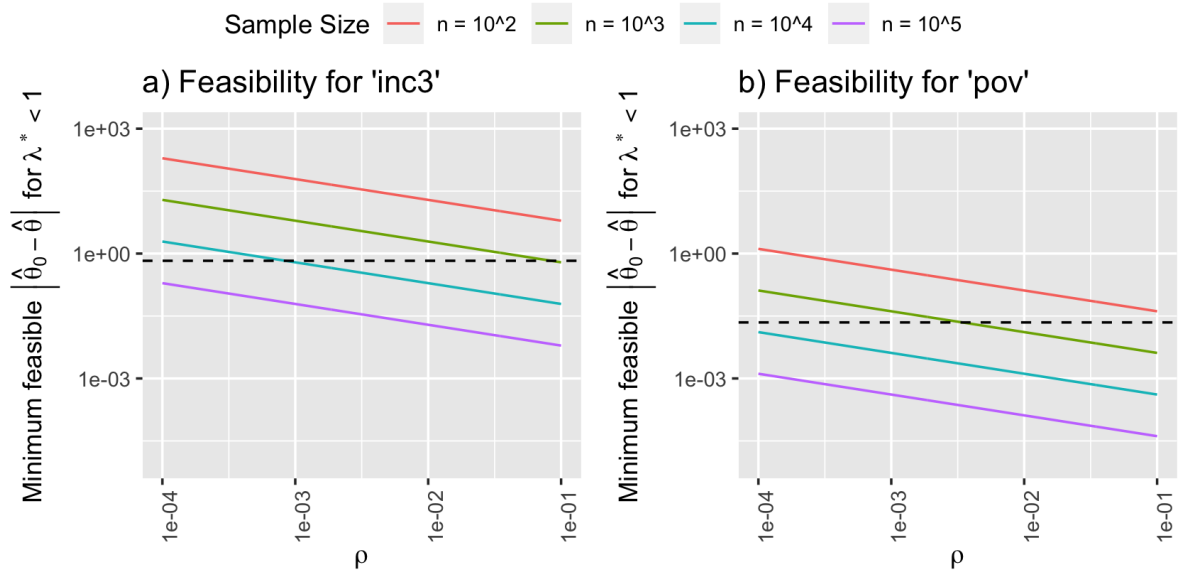



Figure 4: Theoretical minimum AWB for which $\lambda^* < 1$ (y-axis) i.e., survey weighting design is non-ignorable under $\rho$-zCDP, as a function of sample size $n$ (colored lines) and privacy loss budget $\rho$. Subplots and horizontal dashed lines refer to realized AWBs for two variables: `inc3` (left) and `pov` (right).

however, the plug-in estimate improves as $\rho_1$ increases.

### 3.3 End-to-end DP Inferences

In this section, we simulate DP confidence intervals using Algorithms 1 and 2 for the survey weighted population mean of `inc3`, assessing their width and coverage properties. We consider $\rho_1, \rho_2, \rho_3 \in [10^{-3}, 10^{-1}]$, which covers the full spectrum of regularization from $\lambda^*$, as shown in Figure 5. We also vary $\alpha_v$ to show trade-offs between coverage and interval width.

Figure 8 shows boxplots of samples for $\hat{\theta}^{\rho_2\text{-zCDP}}_{\hat{\lambda}^{\rho_1\text{-zCDP}}}$ at different values of $\rho_1$ and $\rho_2$. The red and green dashed lines refer to the unweighted and weighted non-DP estimates, respectively. As $\rho_2$ increases (subplots), we are able to accommodate greater survey weighting corrections relative to the additive noise magnitude. Moreover, for smaller values of $\rho_2$, spending more on $\rho_1$ (colored boxplots) reduces the overall variability of the point estimate.

Figure 9 compares the interval widths of Algorithm 2 to their non-DP counterparts. For each violin plot, we show the distribution of the ratio for the DP confidence interval over the non-DP confidence interval. The dashed horizontal line at 1.0 corresponds to equality. As expected, increasing either $\rho_1$, $\rho_2$, or $\rho_3$ decreases the DP confidence interval width relative to the non-DP interval width. Of particular interest is different values for $\alpha_v$, represented by the different violin plot colors (.5, .05, and .01, respectively). As expected, decreasing $\alpha_v$ gives us wider confidence intervals by accounting for more

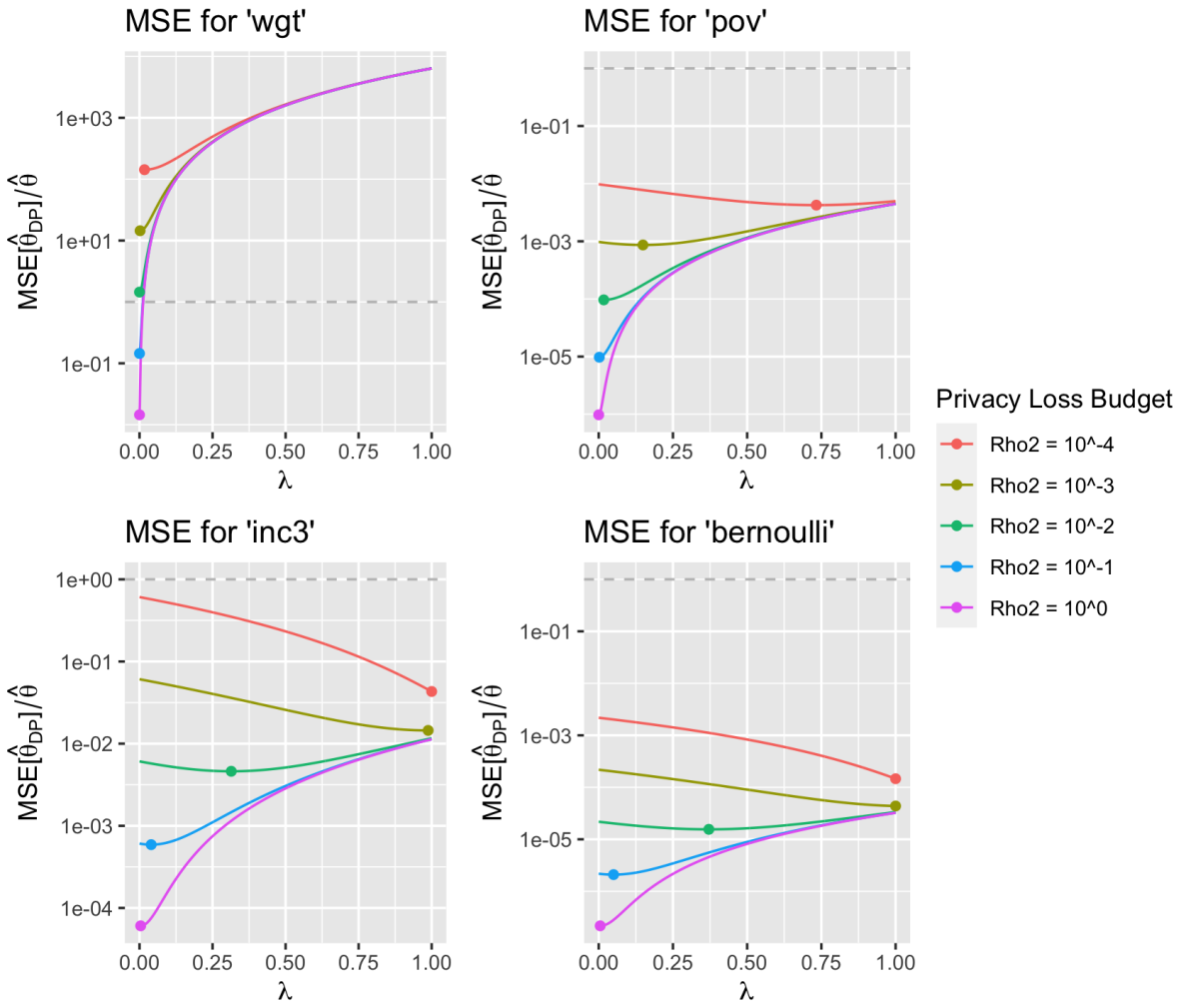



Figure 5: Realized noise-to-signal (DP mean square error divided by non-DP mean estimate, y-axis) as a function of $\lambda$ (x-axis) for different values of privacy loss budget $\rho_2$ (colored lines). Subplots are ordered with decreasing correlation between response variable and survey weights. Points refer to theoretical minimum values, which depend on confidential data and do not satisfy DP.

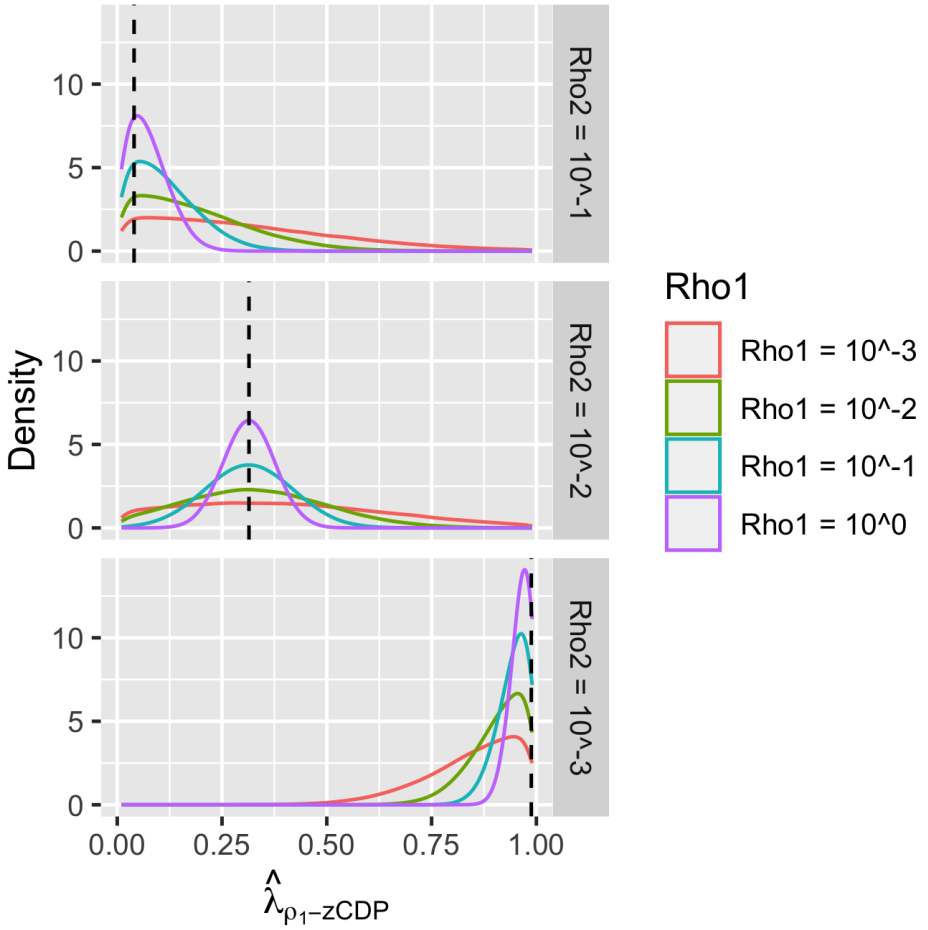



Figure 6: Kernel density estimates for the distribution of $\hat{\lambda}^{\rho_1\text{-zCDP}}$ for different values of $\rho_1$ (colored density plot lines) and $\rho_2$ (subfigures) for `inc3`. Black dashed vertical lines refer to the confidential $\lambda^*$ for each subplot.

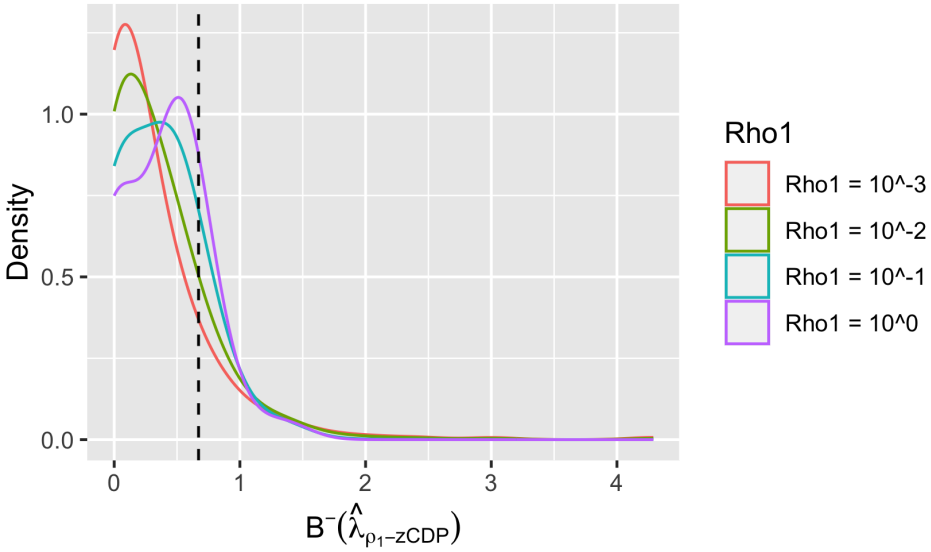



Figure 7: Kernel density estimate for the distribution of the plug-in estimates for AWB for diffferent values of $\rho_1$ (colored density plot lines) for `inc3`. Black dashed vertical line refers to the confidential AWB

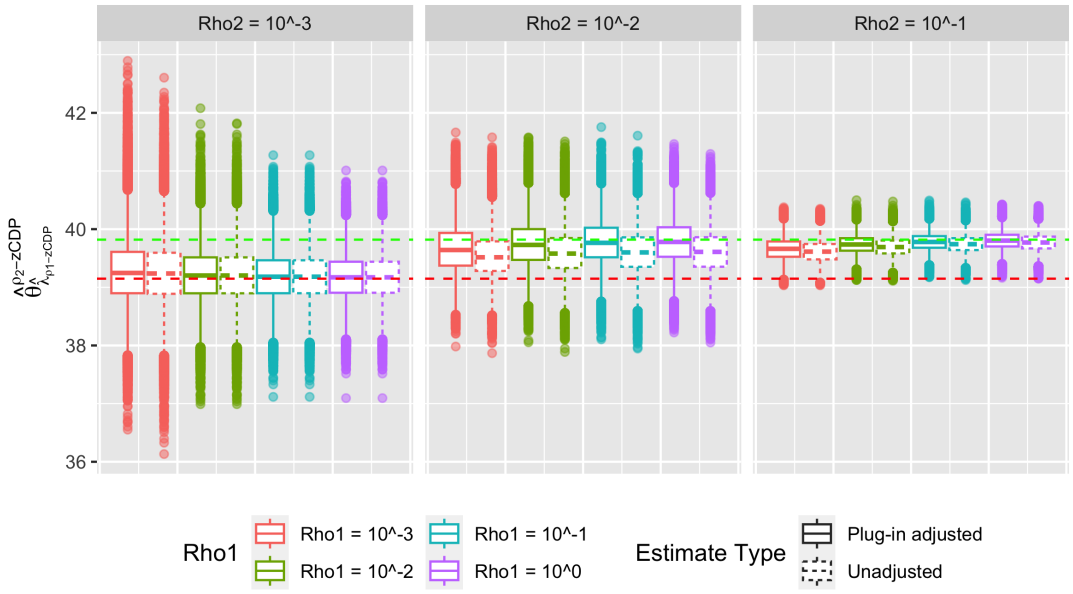



Figure 8: Boxplots of empirical simulations from Algorithm 1 for `inc3` at different values of $\rho_1$ (color) and $\rho_2$ (subfigures). Boxplot outlines refer to the raw estimate (dashed) and the plug-in bias-adjusted estimates (solid). Green dashed line refers to the confidential weighted mean estimate, and the red dashed line refers to the confidential unweighted mean estimate.

potential uncertainty in $\hat{V}$. This greatly improves coverage, which we'll see next.

Figure 10 estimates the average empirical coverage for 95% confidence intervals as a function of $\rho_1, \rho_2, \rho_3$, and $\alpha_v$. To do this, we simulate a "true" parameter $\theta$ from the non-DP normal approximation for each estimate sample, and we empirically average the proportion of intervals covering this simulated ground truth. The dashed line represents 95% coverage, the intended target. As expected, as all the $\rho$ values increase, the empirical coverage tends towards the equivalent non-DP coverage. Similarly, decreasing $\alpha_v$ increases the empirical coverage probability; in particular, when $\rho_3$ decreases, more conservative values of $\alpha_v$ admit better coverage, as expected.

# 4 Discussion

This paper theoretically and empirically suggests that survey weight regularization, when used appropriately can drastically reduces the amount of additional noise needed to preserve DP. By adaptively considering how much to shrink weights while satisfying DP, we produce methods which are operationally feasible while allowing uncertainty quantification at different precisions throughout the entire decision-making process.

While our proposed statistics can admit the construction of valid finite-sample confidence intervals and asymptotic confidence intervals, different methods may produce intervals that are too conservatively wide or liberally narrow in practice. When selecting

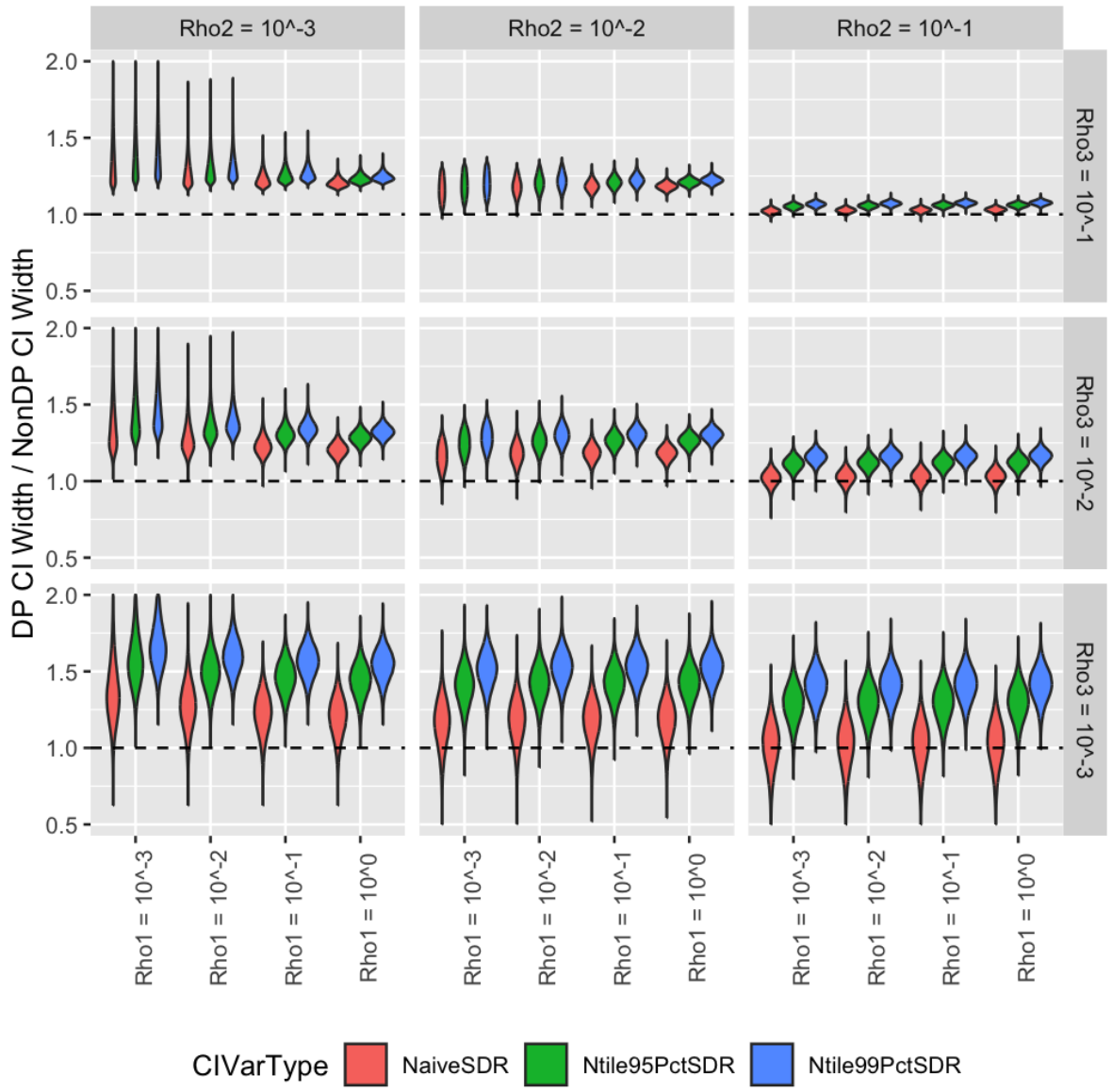



Figure 9: Ratio of DP to non-DP confidence interval widths (y-axis) by values of $\rho_1$ (x-axis), $\rho_2$ (subplot columns), $\rho_3$ (subplot rows), and $\alpha_v$ (colors). Dashed line corresponds to equality (1:1 ratio).

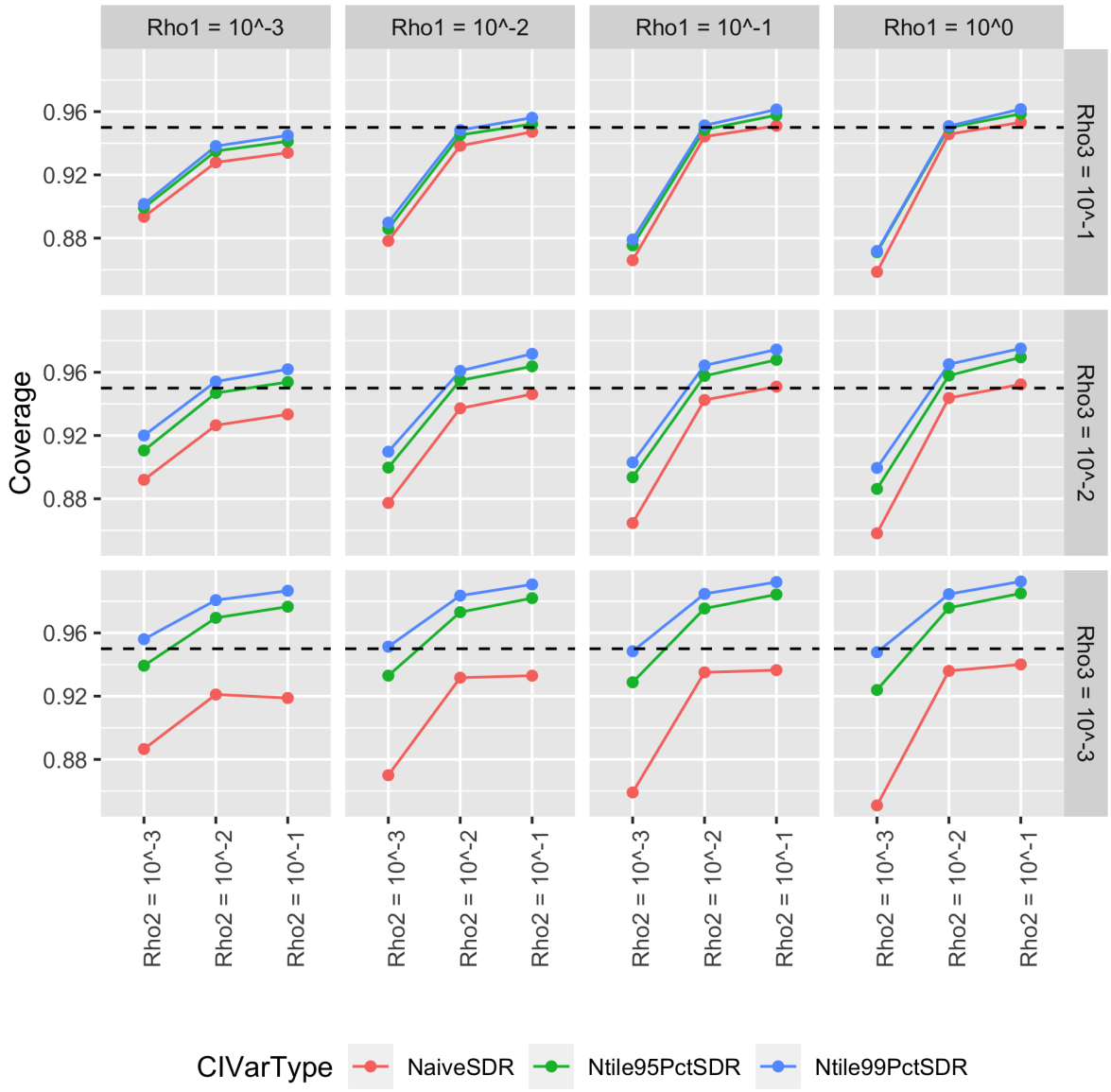



Figure 10: Empirical 95% confidence interval coverage from simulated ground truth population estimates (y-axis) by values of $\rho_1$ (subplot columns), $\rho_2$ (x-axis), $\rho_3$ (subplot rows), and $\alpha_v$ (colors). Dashed line corresponds to 95% coverage, the intended target.

privacy loss budgets for each stage of the algorithm, we recommend incorporating as much domain knowledge as possible. For example, by simulating a distribution of plausible AWB values from prior knowledge, one can establish which kinds of survey weighting biases could be correctable a priori at different privacy loss budgets without peeking at the confidential data.

In future work, we will investigate more complex privacy and utility trade-offs between pure DP and relaxations of DP where not all statistics are subject to the same DP protections. While DP theoretically forbids using data-dependent hyperparameters without DP mechanisms, many commonly used DP algorithms and analyses do not adhere to this rule [1, 25], necessarily yielding additional privacy vulnerabilities in practice [22, 21]. It could be the case that certain hyperparameters could substantially improve the end-to-end usefulness of our estimators at a modest expense to privacy risk, but this would require a much more extensive and nuanced privacy analysis than that offered by a naive comparison of privacy loss budgets. Still, such an analysis could be helpful to understand where DP itself fundamentally limits the kinds of statistical validity offered in survey settings, where worst-case data generating scenarios may be highly unrealistic in practice.

# References


[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

[2] Jordan Awan and Salil Vadhan. Canonical noise distributions and private hypothesis tests. *The Annals of Statistics*, 51(2):547–572, 2023.

[3] Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *Advances in neural information processing systems*, 31, 2018.

[4] Jean-François Beaumont. A new approach to weighting and inference in sample surveys. *Biometrika*, 95(3):539–553, 2008. Publisher: Oxford University Press.

[5] Yves G Berger. Rate of convergence to normal distribution for the horvitz-thompson estimator. *Journal of Statistical Planning and Inference*, 67(2):209–226, 1998.

[6] Kenneth A Bollen, Paul P Biemer, Alan F Karr, Stephen Tueller, and Marcus E Berzofsky. Are survey weights needed? A review of diagnostic tests in regression analysis. *Annual Review of Statistics and Its Application*, 3:375–392, 2016. Publisher: Annual Reviews.

[7] Mark Bun, Jörg Drechsler, Marco Gaboardi, Audra McMillan, and Jayshree Sarathy. Controlling privacy loss in sampling schemes: An analysis of stratified and cluster sampling. In *3rd Symposium on Foundations of Responsible Computing (FORC 2022)*, 2022.

[8] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

[9] Angèle Delevoye and Fredrik Sävje. Consistency of the horvitz–thompson estimator under general sampling and experimental designs. *Journal of Statistical Planning and Inference*, 207:190–197, 2020.

[10] Jörg Drechsler. Differential Privacy for Government Agencies—Are We There Yet? *Journal of the American Statistical Association*, 118(541):761–773, 2023. Publisher: Taylor & Francis.

[11] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[12] Cynthia Dwork, Aaron Roth, and others. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

[13] Andrew Gelman. Struggles with Survey Weighting and Regression Modeling. *Statistical Science*, 22(2):153–164, 2007.

[14] David Haziza and Jean-François Beaumont. Construction of weights in surveys: A review. 2017.

[15] Jingchen Hu, JÖrg Drechsler, and Hang J Kim. Accuracy Gains from Privacy Amplification Through Sampling for Differential Privacy. *Journal of Survey Statistics and Methodology*, 10(3):688–719, 2022. Publisher: Oxford University Press.

[16] Jingchen Hu, Terrance D Savitsky, and Matthew R Williams. Private tabular survey data products through synthetic microdata generation. *Journal of Survey Statistics and Methodology*, 10(3):720–752, 2022. Publisher: Oxford University Press.

[17] Gautam Kamath, Argyris Mouzakis, Matthew Regehr, Vikrant Singhal, Thomas Steinke, and Jonathan Ullman. A Bias-Variance-Privacy Trilemma for Statistical Estimation. *arXiv preprint arXiv:2301.13334*, 2023.

[18] Daniel Kifer, John M Abowd, Robert Ashmead, Ryan Cumings-Menon, Philip Leclerc, Ashwin Machanavajjhala, William Sexton, and Pavel Zhuravlev. Bayesian and Frequentist Semantics for Common Variations of Differential Privacy: Applications to the 2020 Census. *arXiv preprint arXiv:2209.03310*, 2022.

[19] Shurong Lin, Mark Bun, Marco Gaboardi, Eric D Kolaczyk, and Adam Smith. Differentially Private Confidence Intervals for Proportions under Stratified Random Sampling. *arXiv preprint arXiv:2301.08324*, 2023.

[20] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

[21] Shubhankar Mohapatra, Sajin Sasy, Xi He, Gautam Kamath, and Om Thakkar. The role of adaptive optimizers for honest private hyperparameter selection. In *Proceedings of the aaai conference on artificial intelligence*, volume 36, pages 7806–7813, 2022. Issue: 7.

[22] Nicolas Papernot and Thomas Steinke. Hyperparameter tuning with renyi differential privacy. *arXiv preprint arXiv:2110.03620*, 2021.

[23] Aleksandra Slavković and Jeremy Seeman. Statistical data privacy: A song of privacy and utility. *Annual Review of Statistics and Its Application*, 10, 2023. Publisher: Annual Reviews.

[24] Survey Research Center at the Institute for Social Research, University of Michigan Ann Arbor. Panel study of income dynamics, public use dataset, 2019.

[25] Florian Tramer and Dan Boneh. Adversarial training and robustness for multiple perturbations. *Advances in neural information processing systems*, 32, 2019.