An Approximate Monte Carlo Simulation Method for Estimating Uncertainty and Constructing Confidence Intervals for 2020 Census Statistics

Robert Ashmead, Michael B. Hawes, Mary Pritts, Pavel Zhuravlev, and Sallie Ann Keller

April 16, 2024

Abstract

To protect the confidentiality of the 2020 Census, the U.S. Census Bureau adopted a statistical disclosure limitation framework based on the principles of differential privacy. A key component of the 2020 Census Disclosure Avoidance System was the TopDown Algorithm, which applied differentially-private noise to an extensive series of counts from the confidential 2020 Census data and then transformed the resulting noisy measurements into privacy-protected microdata which were then tabulated to produce the 2020 Census Redistricting Data (P.L. 94-171) Summary File as well as the Demographic and Housing Characteristics File. Though the codebase of the implementation, noise distribution, parameters, and noisy measurements are all publicly available, there is not currently a straightforward way to estimate the expected uncertainty for statistics included in the published data products that accounts for both the noisy measurements and the post-processing performed within the TopDown Algorithm. We propose an approximate Monte Carlo Simulation method that allows for the estimation of statistical quantities like mean squared error, bias, and variance, as well as the construction of confidence intervals. The method works by using the output of the *production* iteration of the TopDown Algorithm as the input to additional iterations, allowing for statistical quantities to be estimated from these simulated iterations relative to the original production iteration. This approach is flexible in the sense that it can support the exploration of a wide variety of tabulations and statistical analyses, and it also does not impact the formal privacy protections of the 2020 Census, as it does not directly access any confidential 2020 Census data. In this paper, we empirically test this methodology's validity with a particular focus on creating confidence intervals. The results show that, in general, the quantities estimated by this approximate method closely match their intended targets and that the resulting confidence intervals are statistically valid. Limitations in our method and results are also discussed.

1 Introduction

The application of differentially-private methods in the 2020 Census data products represented one of the largest deployments of a disclosure avoidance system (DAS) based on the principles of formal privacy to date and the first by a national statistical office for one of its primary statistical products. To accomplish the goal of utilizing a differential privacy framework, the U.S. Census Bureau developed the TopDown Algorithm (TDA) [3], which was used to protect the statistics included in the 2020 Census Redistricting Data (P.L. 94-171) Summary File as well as the 2020 Census Demographic and Housing Characteristics File (DHC). The TDA produces privacy-protected microdata: person- and housing unit-level untabulated records. These records are referred to as Privacy-Protected Microdata Files (PPMFs) and are scheduled for public release in 2024.

One of the advantages of utilizing a differentially-private framework is that the methods and parameters of the disclosure limitation methodology can be made fully transparent to the public, meaning that the exact specifications of the algorithms can be released without compromising the formal privacy protections.

By contrast, the integrity of the swapping mechanism used to protect the 2010 Census data products was predicated on keeping the swap rate and other key parameters (and their impacts on data utility) confidential.

One major benefit of the transparency afforded by differentially-private methods is that the information on the random noise distributions utilized as part of the disclosure limitation algorithm can be released to help data users understand the expected level of uncertainty in the published data products. For many differentially-private methods, measures of uncertainty like variance, mean-squared error, bias, and confidence intervals can be directly calculated from the random noise distributions and resulting algorithm outputs. However, this is not the case for statistics protected using the TDA because, while the algorithm uses known random noise distributions to achieve its privacy protections, it subsequently post-processes the noisy measurements along with invariant totals (statistics that are purposely left unchanged, such as state total population counts) and other constraints, to produce person and housing statistics that are non-negative integers that are internally and hierarchically consistent. This consistency ensures that the same statistic calculated from different tables will always yield the same results, such as county totals summing to state totals or counts of males and females summing to the total population count. To achieve these properties, the TDA utilizes numerical optimization methods to solve optimization problems with inequality constraints, such as non-negativity constraints, in its post-processing steps. Consequently, the resulting TDA-protected statistics do not have closed-form formulas for their corresponding measures of uncertainty.

A further consequence of using inequality constraints in the TDA optimization process is that the disclosure avoidance-related variance and bias of the estimators will depend on the underlying data itself rather than solely depending on the randomness of the noisy measurements, which means that applying the algorithm with exactly the same parameters to two different datasets could result in different expected levels of disclosure avoidance-related error. One important category of inequality constraints used in the TDA is the non-negativity constraints. Because the TDA estimate cannot be negative, if a count is zero or close to zero in the underlying confidential data, the TDA estimate will not have the same range of randomness as it would have had if the underlying confidential count was larger. Therefore, the variance will be smaller, and the result may be biased slightly in the positive direction. Similarly, because the TDA kept state total population counts invariant, the small positive bias for small counts can result in a corresponding small negative bias for larger counts.

There is currently no generalized way to estimate the statistical properties of specific queries tabulated from the results of the TopDown Algorithm. The Noisy Measurement Files (NMFs), which contain the noisy measurements observed in the initial stages of TDA iterations, are publicly available for both the Redistricting Data File and the Demographic and Housing Characteristics file 2020 production iterations [2, 1]. These NMFs can be used to generate point estimates and confidence intervals for arbitrary query answers and arbitrary geographic entities composed of block geographic units, but these estimates do not leverage the accuracy improvements provided by the subsequent post-processing steps used by the TDA. The goal of this work is to propose a method that can approximate the statistical properties and uncertainty of any statistic tabulated from the TDA's post-processed privacy-protected microdata output.

The remainder of the paper is organized as follows. In Section 2, we motivate and describe the approximate Monte Carlo simulation method and propose methods to create estimates of statistical quantities and confidence intervals from the simulation iterations. Then, in Sections 3 and 4, we test the methodology on the 2010 Census Redistricting data and Demographic and Housing Characteristics data, respectively. Finally, Section 5 concludes and discusses possible future work.

2   The Approximate Monte Carlo (AMC) simulation method

The primary confidential data input for the Census application of the TDA is the Census Edited File (CEF), the final version of the confidential census data before disclosure avoidance is applied. The CEF consists of records for all housing units and occupied group quarters facilities, and for all persons residing in those housing units and group quarters. For the purposes of estimating the statistical properties of the TDA in this article, we use the CEF as the standard of truth, ignoring other sources of error that may be present in the CEF, such as coverage or measurement error relating to the Census enumeration itself.

Theoretically, we could use a standard Monte Carlo simulation to estimate the statistical properties of the TDA. This would involve running the TDA many times using the same CEF input file, each time taking new

draws of the noisy query measurements from the confidential CEF and independently post-processing those noisy measurements into multiple, independent privacy-protected microdata files. In practice, however, taking additional noisy measurements of the confidential data would contribute to the total privacy loss (multiplying the privacy loss by the number of simulations used), significantly weakening the confidentiality protections provided to census respondents.

The concept behind the proposed *approximate* Monte Carlo (AMC) simulation method is relatively simple. Instead of repeatedly executing the TDA using the confidential CEF data as input, we repeatedly execute the TDA using the already-privacy-protected PPMF *output* of the algorithm as the *input* to the TopDown Algorithm used in the AMC simulations. Because using a realization of the output of the algorithm as the input to further Monte Carlo simulations does not involve accessing the confidential data, the AMC method does not contribute to additional privacy loss.

The motivation here is that if we consider the output of the TDA to be a sufficiently accurate reflection of the CEF itself, then we should get similar statistical properties when using it as the input to the TDA compared with using the CEF as the input. In the limiting case, if it was exactly equivalent to the CEF, then we would get exactly the same distribution. We hypothesize that the less accurate the input to the TopDown Algorithm is relative to the CEF (*e.g.*, the lower the privacy-loss parameter(s) used to generate the noisy measurements from the CEF), the less accurate the AMC simulation method will be.

2.1 Mathematical description

Let θ_{CEF}^p and θ_{CEF}^h be the multidimensional fully saturated contingency tables that are counts of the number of persons and households respectively with specific combinations of characteristics by block in the CEF. θ_{CEF}^p and θ_{CEF}^h represent the data unadjusted for disclosure avoidance. We can therefore think of θ_{CEF}^p and θ_{CEF}^h as the parameters we are estimating in the TDA. Let $TDA(X^p, X^h)$ represent the TDA with a fixed set of parameters relating to the privacy-loss budget and invariants. Let X^p and X^h represent the input person and household datasets. A realization of the TDA applied to the CEF tables is represented as $\{\hat{\theta}^p, \hat{\theta}^h\} = TDA(\theta_{CEF}^p, \theta_{CEF}^h)$ where the dimensionality of $\hat{\theta}^p$ and $\hat{\theta}^h$ will match that of the input data. Let $q()$ represent a generic univariate query (e.g. the number of voting-age persons in Alaska; the number of Hispanic persons in Franklin County, Ohio; the persons per household in a specific census tract) so that $q(\theta_{CEF}^p, \theta_{CEF}^h)$ would represent the answer to the query using the CEF persons and/or housing data.

2.1.1 Standard Monte Carlo estimates

Consider a number (m) of independent realizations of the TopDown Algorithm from the CEF data

$$\{\hat{\theta}_i^p, \hat{\theta}_i^h\} \sim TDA(\theta_{CEF}^p, \theta_{CEF}^h), \ i = 1, \cdots, m.$$

Monte Carlo empirical estimates of bias, variance, and MSE for any query $q(\cdot)$ could be calculated as

$$\widehat{BIAS}_{MC} = \frac{1}{m} \sum_{i=1}^{m} q(\hat{\theta}_i^p, \hat{\theta}_i^h) - q(\theta_{CEF}^p, \theta_{CEF}^h) \tag{1}$$

$$\widehat{VAR}_{MC} = \frac{1}{m-1} \sum_{i=1}^{m} \left(q(\hat{\theta}_i^p, \hat{\theta}_i^h) - \overline{q(\hat{\theta}_i^p, \hat{\theta}_i^h)} \right)^2, \tag{2}$$

where $\overline{q(\hat{\theta}_i^p, \hat{\theta}_i^h)} = \frac{1}{m} \sum_{i=1}^{m} q(\hat{\theta}_i^p, \hat{\theta}_i^h)$, and

$$\widehat{MSE}_{MC} = \frac{1}{m} \sum_{i=1}^{m} \left(q(\hat{\theta}_i^p, \hat{\theta}_i^h) - q(\theta_{CEF}^p, \theta_{CEF}^h) \right)^2, \tag{3}$$

respectively. As m goes to infinity, assuming finite variance and appropriate finite moments, these calculations will be asymptotically equal to the bias, variance, and MSE of the query for the algorithm applied to the CEF.

2.1.2 Approximate Monte Carlo (AMC) simulation estimates

Let $\{\hat{\theta}_0^p, \hat{\theta}_0^h\} = TDA(\theta_{CEF}^p, \theta_{CEF}^h)$ be a single realization from the TopDown Algorithm that represents the production iteration.[1] Next, assume we simulate s independent realizations of the algorithm using $\{\hat{\theta}_0^p, \hat{\theta}_0^h\}$ in place of the CEF as input into the TDA, meaning we have

$$\{\tilde{\theta}_i^p, \tilde{\theta}_i^h\} \sim TDA(\{\hat{\theta}_0^p, \hat{\theta}_0^h\}), \ i = 1, \cdots, s.$$

Then, Monte Carlo empirical estimates of bias, variance, and MSE for any query $q(\cdot)$ *relative to* $\hat{\theta}_0^p, \hat{\theta}_0^h$ could be calculated as, respectively

$$\widehat{BIAS}_{AMC} = \frac{1}{s} \sum_{i=1}^{s} q(\tilde{\theta}_i^p, \tilde{\theta}_i^h) - q(\hat{\theta}_0^p, \hat{\theta}_0^h), \tag{4}$$

$$\widehat{VAR}_{AMC} = \frac{1}{s-1} \sum_{i=1}^{s} \left(q(\tilde{\theta}_i^p, \tilde{\theta}_i^h) - \overline{q(\tilde{\theta}_i^p, \tilde{\theta}_i^h)} \right)^2, \tag{5}$$

where $\overline{q(\tilde{\theta}_i^p, \tilde{\theta}_i^h)} = \frac{1}{s} \sum_{i=1}^{s} q(\tilde{\theta}_i^p, \tilde{\theta}_i^h)$, and

$$\widehat{MSE}_{AMC} = \frac{1}{s} \sum_{i=1}^{s} \left(q(\tilde{\theta}_i^p, \tilde{\theta}_i^h) - q(\hat{\theta}_0^p, \hat{\theta}_0^h) \right)^2. \tag{6}$$

The proposed AMC simulation method relies on the assumption that for a given query of interest, $BIAS_{AMC} \approx BIAS_{MC}$, $VAR_{AMC} \approx VAR_{MC}$, and $MSE_{AMC} \approx MSE_{MC}$ when $\{\theta_{CEF}^p, \theta_{CEF}^h\} \approx \{\hat{\theta}_0^p, \hat{\theta}_0^h\}$. If those quantities are similar, we can utilize the approximate quantities for describing the statistical properties of the algorithm and constructing confidence intervals, all without adding to the privacy loss.

2.2 Constructing confidence intervals from the simulations

For a given query $q(\cdot)$, we want to construct a confidence interval that reflects the uncertainty of $q(\hat{\theta}_0^p, \hat{\theta}_0^h)$, the query answer from the production iteration of the algorithm. We examine two basic strategies for constructing confidence intervals from the approximate Monte Carlo simulations. The first is a Wald-type interval, assuming a Gaussian or Student's t distribution for the error. Given that the noisy measurements from the TDA were generated using a discrete Gaussian distribution, this would seem to be a reasonable assumption, though one we will examine further. The second is a quantile-based confidence interval based on empirical quantiles of the approximate Monte Carlo simulations. Additionally, we propose and examine bias-adjusted confidence intervals to better account for bias in the TDA estimates as measured by the AMC simulations.

2.2.1 Wald-type confidence intervals

If we were to assume that the query estimates produced by the TDA are unbiased and approximately normally distributed, then a Wald-type interval given by

$$q(\hat{\theta}_0^p, \hat{\theta}_0^h) \pm \mathcal{Z}_{1-\frac{\alpha}{2}} \sqrt{\widehat{VAR}_{AMC}} \tag{7}$$

would be the best choice. As acknowledged above, however, we know that in some cases the TDA yields biased query estimates which will decrease the accuracy and coverage of our confidence intervals. Replacing the estimated variance with the estimated mean-squared error results in a more statistically conservative interval. When the estimated bias is small, the width of the interval will be the same or only slightly larger as if we had used the variance. For larger estimated bias, the interval width will increase proportionally. That confidence interval is given by

$$q(\hat{\theta}_0^p, \hat{\theta}_0^h) \pm \mathcal{Z}_{1-\frac{\alpha}{2}} \sqrt{\widehat{MSE}_{AMC}}. \tag{8}$$

---

[1]This is comparable to the single realization that is used to create the PPMF and associated published tabulations.

In addition to the Gaussian distributions for the margins-of-error, we also consider whether a heavier-tailed distribution like a Student's t distribution is more appropriate. We chose to use 5 degrees based on examining the results of our experiments;[2] however, we don't have a theoretical rationale for this value. Lastly, if the query of interest is a count, we propose to apply a floor and ceiling function to the endpoints of the confidence intervals respectively and truncate at zero to account for the outcome space being integer and non-negative.

2.2.2 Quantile-based confidence intervals

If we have a large enough number of simulations, another strategy for constructing confidence intervals would be to use a non-parametric method based on the quantiles of the simulated query answers. Let $p_\alpha(q(\{\tilde{\theta}_i^p, \tilde{\theta}_i^h\}))$ represent the αth percentile of the approximate Monte Carlo Values from the simulated estimates of the quantile. Then, for example, a 90% approximate confidence interval could be estimated as

$$\left(p_{05}(q(\tilde{\theta}_i^p, \tilde{\theta}_i^h)), p_{95}(q(\tilde{\theta}_i^p, \tilde{\theta}_i^h)) \right) \tag{9}$$

If the query of interest is a count, we propose to apply a floor and ceiling function to the endpoints of the confidence intervals, respectively, to account for the outcome space being integer. A major downside of the quantile-based method is that in order to calculate relatively precise quantiles, it is necessary to have a large number of simulations. Given that the TDA is computationally expensive to run[3], this is a major limitation.

2.2.3 Bias-corrected confidence intervals

We know that in some cases, the TDA estimator is biased. Therefore, if the estimate of bias from the AMC method is accurate, then an improved point estimate and confidence interval could be constructed for the Wald-type intervals by subtracting the estimated bias from the point estimate and using that as the pivotal value for the confidence interval. For quantile-based intervals, we propose subtracting the estimated bias from each of the endpoints of the confidence interval.

2.3 Comparisons with existing literature

Several authors have proposed bootstrap-related methods which are similar to the proposed AMC method in that they attempt to estimate the statistical properties of a differentially-private estimator empirically rather than using closed-form formulas based on the noise distribution. Brawner and Honaker proposed a bootstrap method in which the original dataset is bootstrapped before noisy queries are generated and then the average and standard deviation of the bootstrapped answers is used to construct a confidence interval [5]. Others have extended these ideas [12, 7]. Relatedly, parametric bootstrap methods [9, 11] have also been proposed, which apply when a distributional form has been assumed for the underlying data.

Preliminary work on the AMC simulation method proposed in this article applied an early version of the TDA to the 1940 Decennial Census and found that the method did well in estimating variance [4]. While our methods share some of the same ideas and motivations as the bootstrap methods, our specific experiment setup is unique in several important ways. First, the TDA is designed to produce microdata. As a consequence, the input and output parameter space of the algorithm are equivalent, allowing us to use the algorithm output as the input to our simulations. Second, many of the bootstrap methods rely on incorporating the methodology as part of the design of the differentially private algorithm. In the case of the census use case, the methodology for creating confidence intervals was applied afterwards. Lastly, the TDA is computationally expensive, so it is necessary to keep the number of simulations or bootstraps to a modest number.

---

[2]The critical value for a 90% confidence would be 2.015 vs. 1.645 for the Gaussian distribution, an approximately 25% increase.

[3]A single run of the TDA will generally cost around $1,500 for the redistricting data schema, and around $10,000 for the DHC data schema

3 Experimental results using 2010 Census Redistricting data

In order to test the validity of the AMC method's application to the TDA, we utilized Monte Carlo simulations of the redistricting production settings on the 2010 Decennial Census data. First, we completed 100 independent simulations of the TDA using the confidential 2010 CEF as input. These simulations represent the sampling distribution of the TopDown Algorithm applied to the CEF and serve as a gold standard to the AMC method. Second, using a single privatized PPMF output[4] (which we call $PPMF_0$) from the first set of iterations in place of the CEF, we completed 100 additional simulations of the algorithm to be used for the AMC estimates.

In order to test the methodology, we selected all queries (301 in total[5]) from the redistricting data summary file and tabulated them using the simulations at multiple geographic levels. Queries were taken from the following tables:

- P1. Race

- P2. Hispanic or Latino, and not Hispanic or Latino by Race

- P3. Race for the Population 18 Years and Over

- P4. Hispanic or Latino, and not Hispanic or Latino by Race for the Population 18 Years and Over

- P5. Group Quarters Population by Major Group Quarters Type

- H1. Occupancy Status (Housing)

Given the large number of geographies at the block level, we selected a stratified sample of blocks to keep the experiment analysis more manageable. The strata were created by grouping blocks by CEF total population. We randomly sampled 5 blocks for each state within each strata.

This experiment has three major goals. The first is to determine how well the approximate quantities match their empirical counterparts. We do this by comparing the empirical RMSE, bias, and standard deviation estimates graphically. The second is to determine if confidence intervals based on the approximate quantities have the appropriate coverage. In order to properly test the coverage, it would be necessary to repeat the entire simulation many times starting with creating a new $PPMF_0$ followed by the AMC simulations. This is too computationally expensive for this use case, so instead we consider the percent of confidence intervals that contain the CEF query value across different queries. Lastly, we examine if it is feasible to utilize the method with a fewer number of AMC simulations.

3.1 Results comparing RMSE, bias, and standard deviation estimates

We created graphics to compare RMSE, bias, and standard deviation estimates for each of the three statistics and for queries within each of the following geographic levels: U.S., state, county, tract, block, American Indian/Alaska Native (AIAN) area, and elementary school districts. Complete results are available in the online appendix. Figures shown here are meant to be representative of larger patterns or to communicate a specific point.

Figure 1 shows a set of hexplots of the RMSE of the $15,652$ query estimates[6] at the state level with the estimate coming from the CEF-generated calculations on the x-axis and the $PPMF_0$-generated calculations along the y-axis. Hexplots divide the plotting area into hexagons each reflecting locations and density of the $15,652$ points. The darker the shade, the more individual points were in the hexagon. The plot is faceted by the CEF value of the query and note that the range of the axes changes between subplots. Ideally, we would like to see all the points fall along the diagonal which would indicate that the AMC-based values are

---

[4]Technically the output of the TopDown Algorithm is called the microdata detail file (MDF), the format and metadata of which differ slightly from published PPMFs. We used the MDF for our experiments, but all of the necessary detail is available on the PPMF.

[5]2020 Census Redistricting (P.L. 94-171) Data Summary File Technical Documentation, available at: `https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_StatesTechDoc_English.pdf`

[6]301 queries x 52 geographies
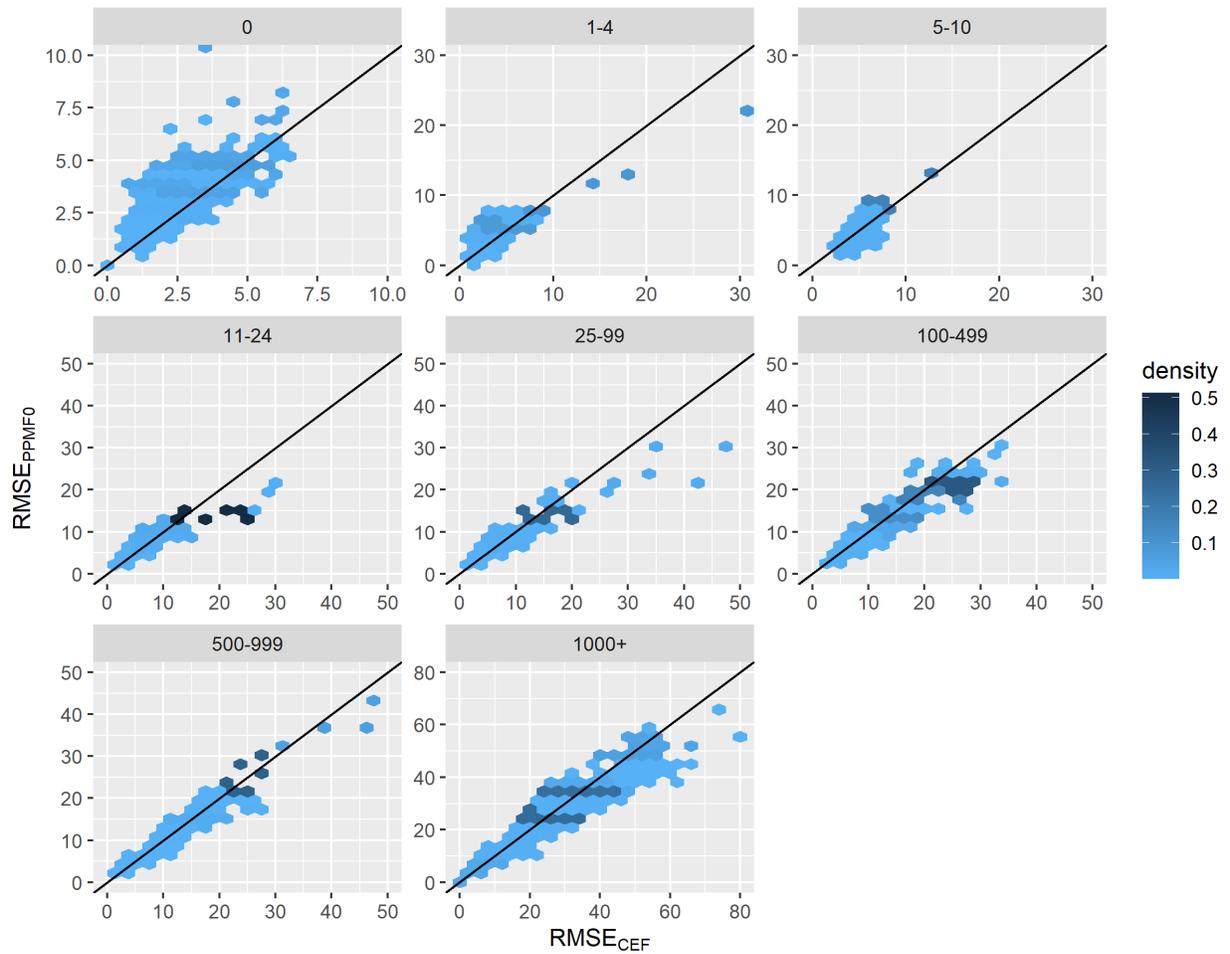
Figure 1: State Level estimated $RSME_{PPMF0}$ from $PPMF_0$ vs estimated $RMSE_{CEF}$ from the CEF is shown. Each panel correspond to certain sizes (i.e. counts) of the Query (based on CEF). In case when RMSEs-from-PPMF0 are very similar to RMSEs-from-CEF they line up along the diagonal, showing the AMC to be a good approximation. This is mostly the case with the data shown. (Hexagons show location and density of the $15,652$ (301 queries x 52 geographies) points used for the plot.

7

equivalent to their CEF-based counterparts and could therefore be substituted for one another freely. For the most part, we see that the estimates closely match one another. Most of the data points are close to the diagonal line; however, in the subplots for CEF query size 0 and 1-4, we see some points were above the diagonal line, indicating the RMSE from the AMC was slightly larger than from the CEF in some cases.

Looking at the bias from the standard Monte Carlo method (which used the CEF as input to the TDA) compared to the bias from the AMC method, we saw the estimates generally match one another across geographies and statistics; however, as we get to smaller geographies (tract, block, and elementary school district) we start to see some underestimation of the bias from the AMC method when it is further from zero. See for example Figure 2. This is most apparent for queries in the larger groupings. As a result, the RMSE estimates of those queries for the AMC are also slightly underestimated.
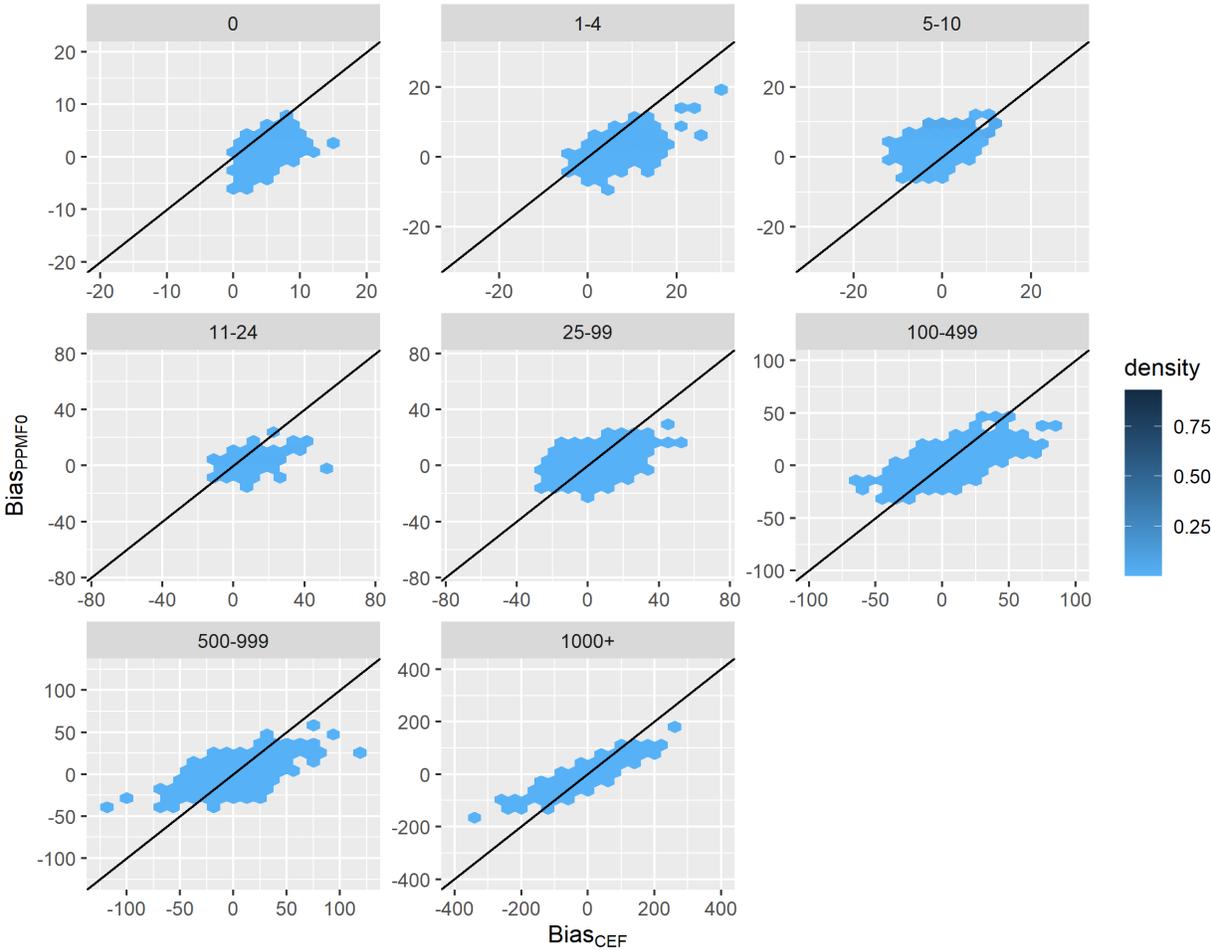
Figure 2: Elementary school district level estimated bias from the $PPMF_0$ vs estimated bias from the CEF, grouped by CEF-based query size. Some underestimation of the bias in the AMC method can be seen (which leads to underestimating RMSE as well).

3.2 Are TDA estimates approximately normally distributed?

In order for the Wald-type confidence intervals to be valid, the distribution of the TDA estimates for a query needs to be approximately normally distributed. We examine this assumption by looking at the empirical distributions of 100 TDA runs on four selected queries in Figure 3 through a quantile-quantile plot, which compares the two distributions to each other. Queries 1, 3, and 4 have distributions that very closely resemble
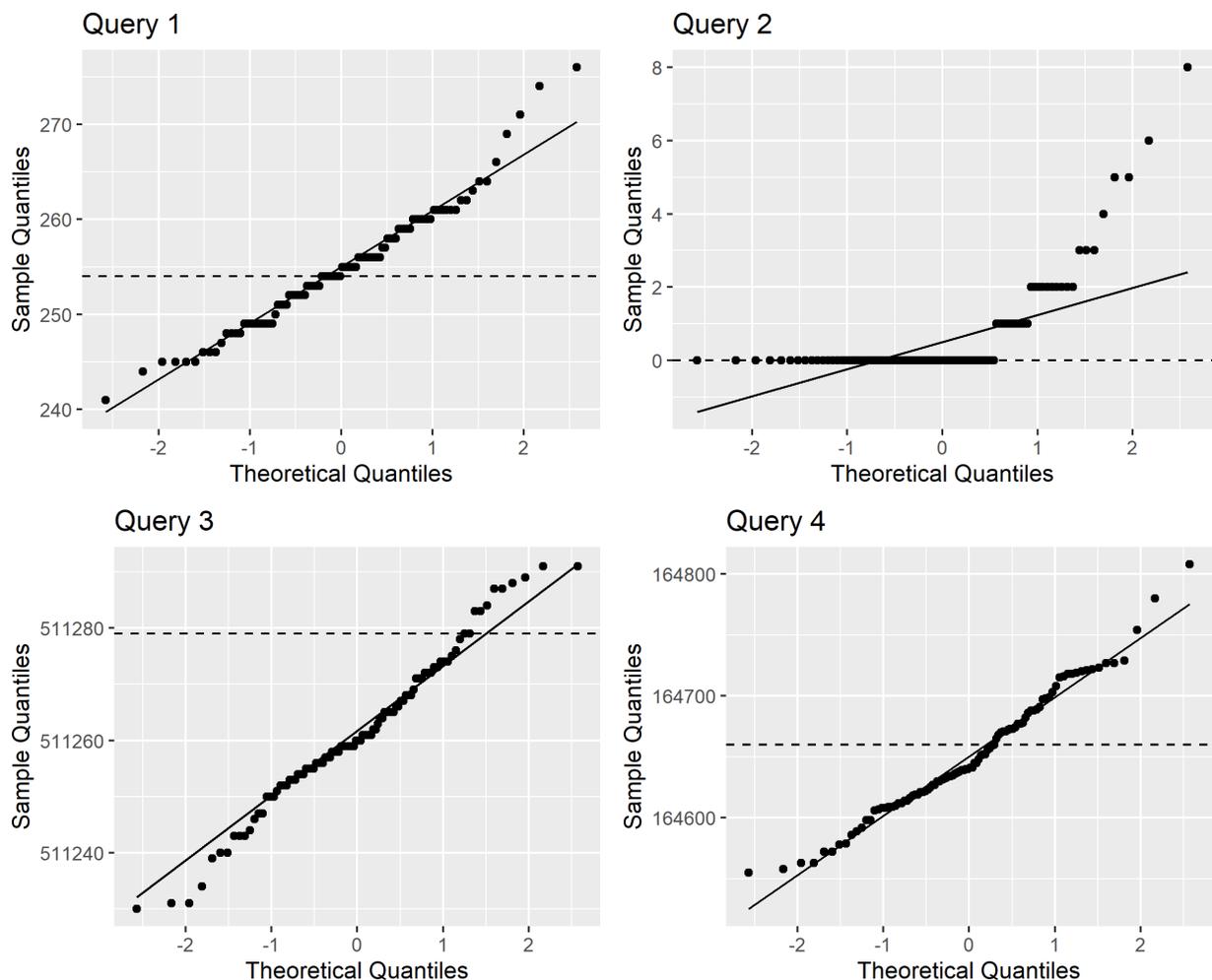
Figure 3: Quantile-quantile plot for four 2010 redistricting queries at the state level over 100 AMC iterations visualize the validity for the Wald-type confidence intervals. Query 1 is the P0040020 query (Population of Two Races: Black or African American; Native Hawaiian and Other Pacific Islander) for Arizona; Query 2 is the P0040064 query (Population of Four Races: American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race) for New Jersey; Query 3 is the P0010003 query (Population of One Race: White Alone) for Wyoming; Query 4 is the H0010003 query (Vacant Units) for Maine. For each query, there are 100 query counts from the AMC iterations plotted against the quantiles of the theoretical normal distribution. Horizontal dashed lines represent $PPMF_0$ value. Theoretical quantiles are those of a Gaussian distribution using the estimated mean and variance from the 100 iterations. Sample quantiles for Queries 1, 3 and 4 correspond very closely to theoretical normal distribution quantiles (less closely at the tails). Query 2 had $PPMF_0$ value (count) of zero, and non-negativity constraints rendered the distribution obviously non-gaussian.

a normal distribution. In each of those cases, there are some points on the tail that are slightly more extreme than the theoretical quantiles. The TDA estimates from Query 3 are also biased in that the $PPMF_0$ value is about 20 larger than the center of the distribution. Query 2 is a case where the TDA distribution is obviously not normally distributed because of restriction to non-negative values. In this case the $PPMF_0$ value of the query was zero and most of the resulting TDA estimates were also zero, but a few end up being larger. While the results from queries like Query 2 are clearly not normally distributed, we'll see later on from our experiments that the confidence intervals still appear to be valid (and actually statistically more conservative than for larger queries) in terms of their coverage. These queries are representative of what we found in our

exploratory results and while these examples are in no way exhaustive of the possible queries, they show that, for queries that are substantially larger than zero, a normal approximation or a similar distribution with slightly thicker tails is a reasonable assumption to use when constructing a confidence interval about the TDA estimates.

3.3 Confidence intervals for Redistricting data

A valid 90% confidence interval estimator will contain the true population parameter at least 90% of the time upon repeated sampling. In an ideal experiment we could calculate the coverage for each individual query across multiple iterations of the process. In this scenario, that would include multiple iterations of both the $PPMF_0$ creation and the subsequent additional AMC iterations. This was judged to be infeasible given the computational cost of running the full TDA at scale. Instead, we compute the proportion of the confidence intervals that include the population parameter (here the confidential CEF-based count) for different groups of queries and for a single iteration $PPMF_0$. This should not be considered the actual coverage of the confidence intervals because confidence intervals for different queries are correlated to varying degrees; however, we do think it is still adequate in assessing the validity of the methodology given that many queries (especially across geographies and about different attributes) should have small correlation.

In this experiment, we considered eight different types of nominal 90% confidence intervals.[7] Labels for the CI type are in parentheses (see Table 1 for reference). First, we considered the non-parametric quantile-based method (np) and a bias-corrected non-parametric quantile-based method (BCnp) which subtracted the estimated bias (constructed from the difference between the median value and the $PPMF_0$ value) from the interval. Next, we constructed a RMSE-based confidence interval assuming a normal distribution like Equation 8 (z) and a Student's t distribution with 5 degrees of freedom (t). Confidence interval types (BCz) and (BCt) are bias-corrected versions of z and t intervals in which the estimated bias was subtracted from the interval. Lastly, versions of the confidence intervals were constructed using a conditional approach in which the bias-corrected interval was used only if some criteria was met and otherwise the non-bias-corrected versions were utilized. These were created for both the Gaussian (cz) and Student's t (ct) based CIs.

The criteria for using the bias-corrected versions were

- The $PPMF_0$ value was greater than 5; and

- The absolute value of the estimated bias divided by the estimated standard deviation of the AMC estimates was greater than or equal to 0.5; and

- The estimated bias was either negative or the $PPMF_0$ value was at least 25.

The rules for the conditional intervals were determined based on the observations that the estimated bias for very small counts was often not very reliable and that the estimated bias for small counts was almost always positive. The rule concerning the size of the bias relative to the standard deviation was motivated by the desire to only make the adjustment when the estimated bias would be impactful. The thresholds for the three criteria were selected based on our empirical testing. With the conditional rules, the majority of the time the criteria was not met, but when it was it typically made the confidence intervals more accurate.

Label	Description
np	Non-parametric quantile-based CI
BCnp	Bias-corrected non-parametric quantile-based CI
z	RMSE-based Wald CI with normal distribution
t	RMSE-based Wald CI with Student's t distribution
BCz	Bias-corrected RMSE-based Wald CI with normal distribution
BCt	Bias-corrected RMSE-based Wald CI with Student's t distribution
cz	Conditionally bias-corrected RMSE-based Wald CI with normal distribution
ct	Conditionally bias-corrected RMSE-based Wald CI with Student's t distribution

Table 1: Confidence Interval Types Used in Experiments

[7]90% confidence intervals are the standard of the U.S. Census Bureau

Figures 4 and 5 show the proportions of the confidence intervals that include the CEF values by query size (count) groupings for the AIAN area and block-level queries, respectively. The AIAN area results are similar to those of many of the larger geographies, while the block-level results show the benefit of the bias-correction for specific cases. Results for additional geographies are shown in the online appendix. Across all the geographies, the two non-parametric methods (np and BCnp) were the least likely to meet the 0.9 threshold, especially for larger queries. Conversely, the Wald-type intervals were consistently able to meet or surpass the 0.9 threshold. The intervals using the Student's t quantile (t, BCt, ct) are wider and thus have slightly higher coverage than the (z, BCz, cz) similar interval using a normal quantile. For the Wald-type intervals, the bias-correction mostly does not make a difference and sometimes slightly decreases the proportion of the CIs that contain the CEF value. In a few cases, like the two largest query size (count) groups in Figure 5, the bias-correction makes a positive and important difference. The conditional bias-corrected confidence intervals (cz, ct) do a fairly good job of balancing the best properties of the unadjusted and bias-adjusted intervals.
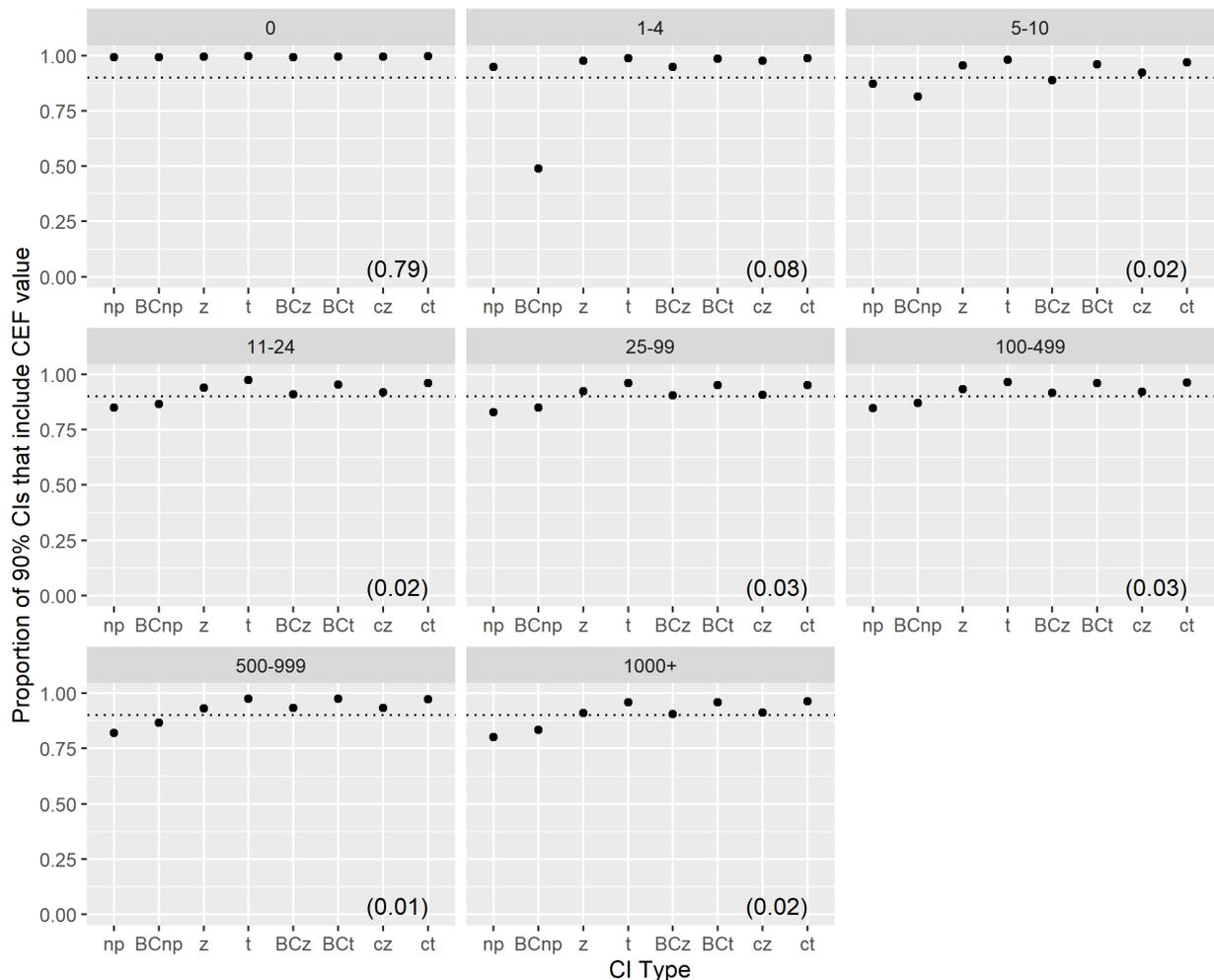
Figure 4: The proportion of 90% CIs that contained the CEF value for 2010 AIAN area level PL94 queries aggregated by query count (size) (the bracketed number in the lower right corner indicates the proportion of the queries in that size category, e.g. in this figure most – 0.79 – have value of 0). The Wald-type intervals all do similarly well and mostly surpass 0.9 threshold. Student's t based CIs (t, BCt, ct) are wider and have slightly higher coverage. The most poorly performing are the non-parametric ones (np and BCnp). Not much difference was made by the bias correction.
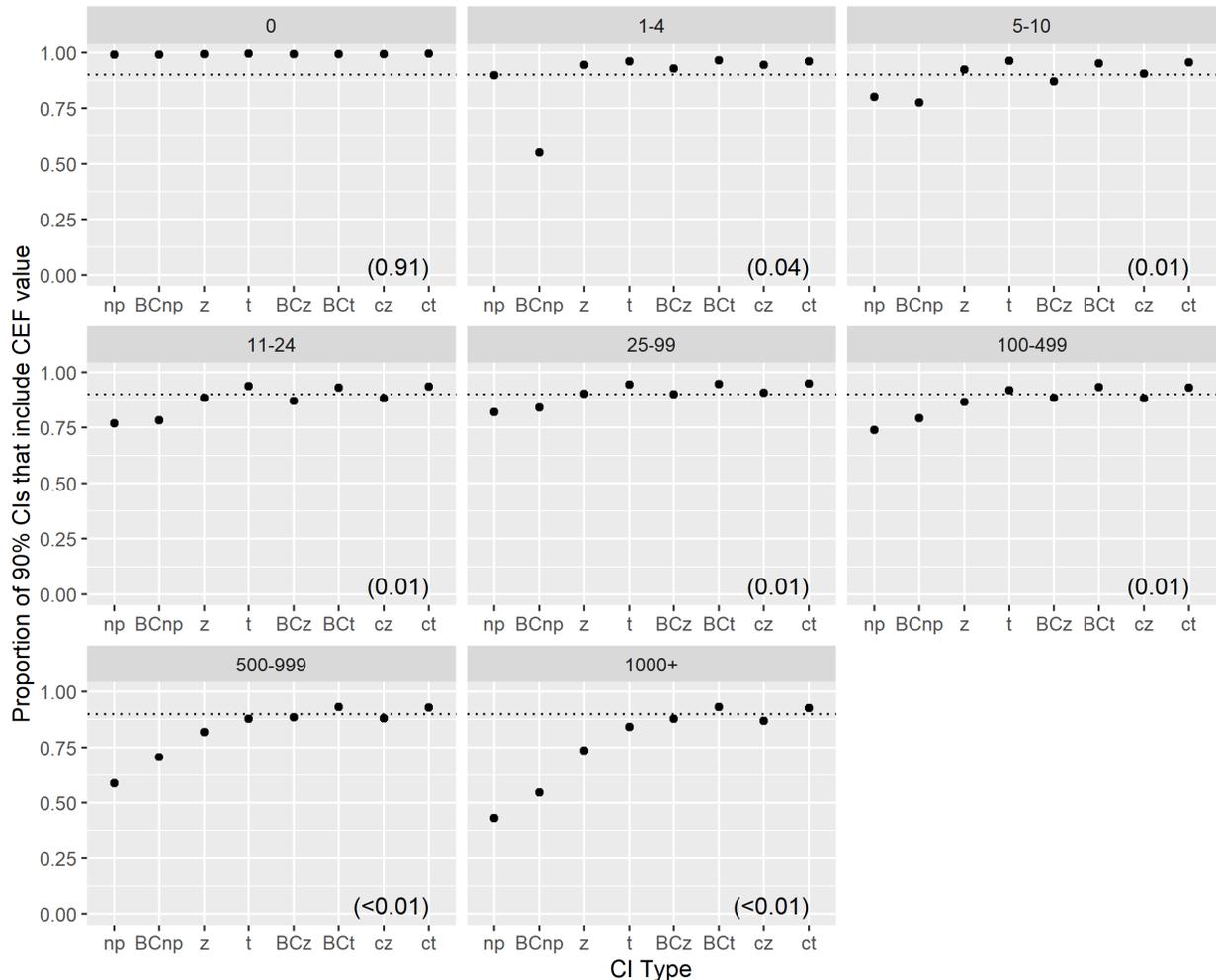
Figure 5: The proportion of 90% CIs that contained the CEF value for 2010 block level PL94 queries aggregated by query size (the bracketed number in the lower right corner indicates the proportion of the queries in that size category, e.g. in this figure most – 0.91 – have value of 0). Non-parametric CIs perform the most poorly with respect to containing the CEF value. For the Wald-based CIs, the ones using Student's t distribution performed slightly better than the ones using the normal distribution. Bias correction slightly improves performance only in the last two panels (largest query size (count) groups). Conditional bias correction (ct) combines the best properties of bias-uncorrected and bias-corrected CIs.

3.4   Confidence interval width

We compared summaries of confidence interval widths between the different methods across multiple geographic levels. The summary for the county level queries is shown in Figure 6 and is generally representative of the patterns across CI type and query size. Additional box plots are shown in the online appendix. We omit the bias-adjusted non-parametric and conditional confidence interval methods in these plots as the widths are similar to the results shown (BCnp is similar to np; ct and cz were similar to z and t, respectively). The widths were generally slightly smaller for the non-parametric CIs which relates to the smaller probability of containing the CEF value that we observed in Figures 4 and 5. As we would expect, t and BCt intervals are slightly wider than the z and BCz intervals. In general, the confidence interval width increases as the size of the query increases. The distributions of widths are relatively similar across geographic levels within the query size groupings. In general, the CI widths appear to be small enough to provide meaningful

information to users, with the exception of some outliers.
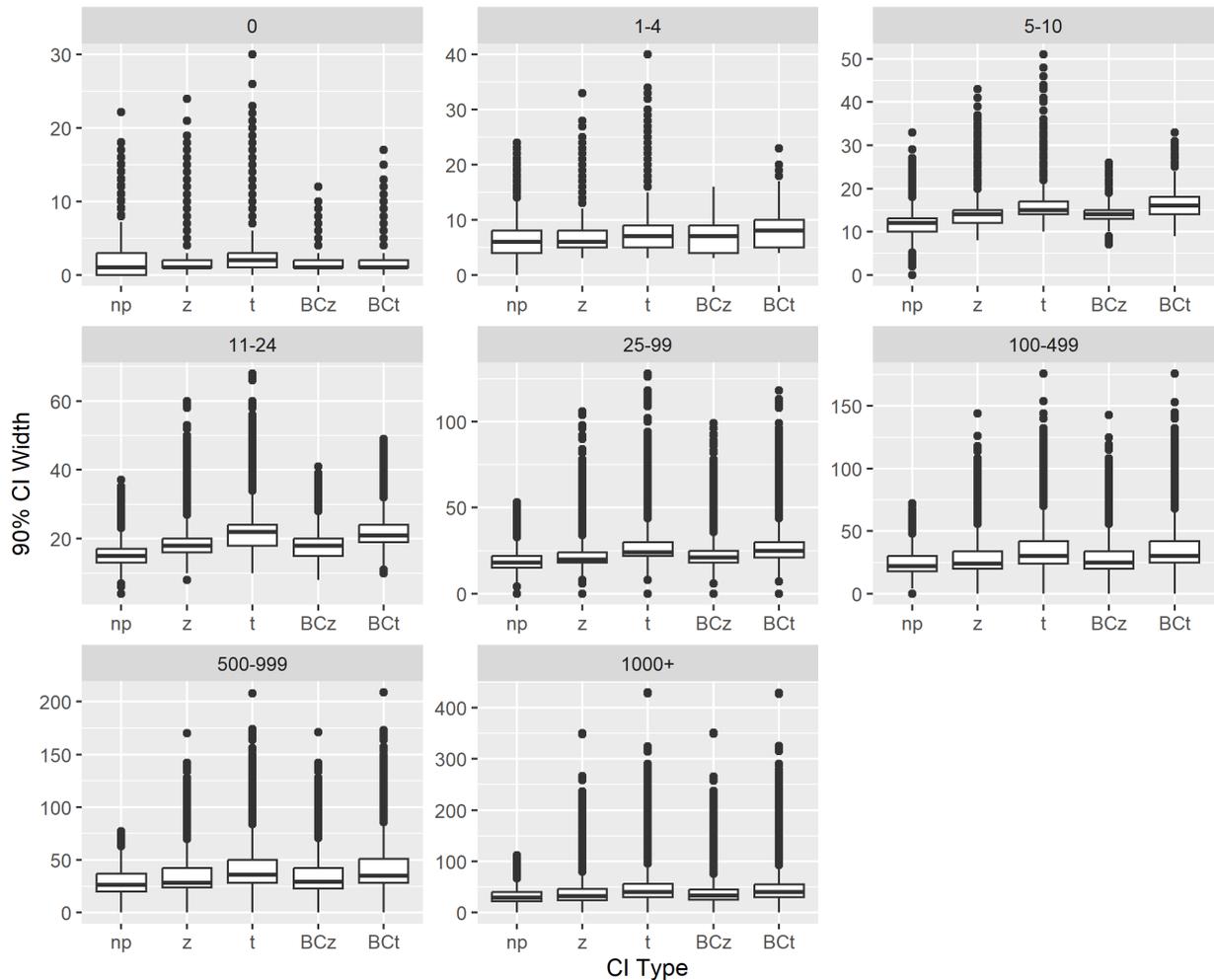
Figure 6: The distribution of 90% CIs widths for 2010 County Level PL94 queries aggregated by query size. Expectedly, t and BCt intervals are slightly wider than the z and BCz intervals; also the confidence interval width generally increases as the size of the query increases.

3.5   The Effect of the number of AMC simulation iterations on the estimates

Running the TDA incurs a significant computational cost and being able to reduce the number of iterations needed to estimate statistical quantities used for the confidence intervals is beneficial. Non-parametric quantile methods were not considered in the remaining analysis because they did not perform sufficiently well based on 100 iterations. In order to test whether fewer iterations would be feasible using the Wald-type intervals, we used random subsets of sizes 25, 50, and 75 of the 100 iterations and examined the difference in the proportion of the confidence intervals that contained the CEF value and how the estimated statistical quantities (bias, RMSE) compared to one another. We found that reducing the iterations to 25 yielded RMSE metrics that were only slightly less precise and only slightly lower on average. Figure 7 shows the proportions of ct 90% confidence intervals that contained the CEF value by geographic level and query grouping. Reducing to 25 iterations decreased the proportion slightly in most of the groupings compared with using a higher number of iterations, but still generated intervals that met the 90% threshold. Given these results, we decided to use only 25 Monte Carlo iterations in our additional experiments below that

explore applying our proposed approach to the Demographic and Housing Characteristics PPMFs that are output from the TDA.
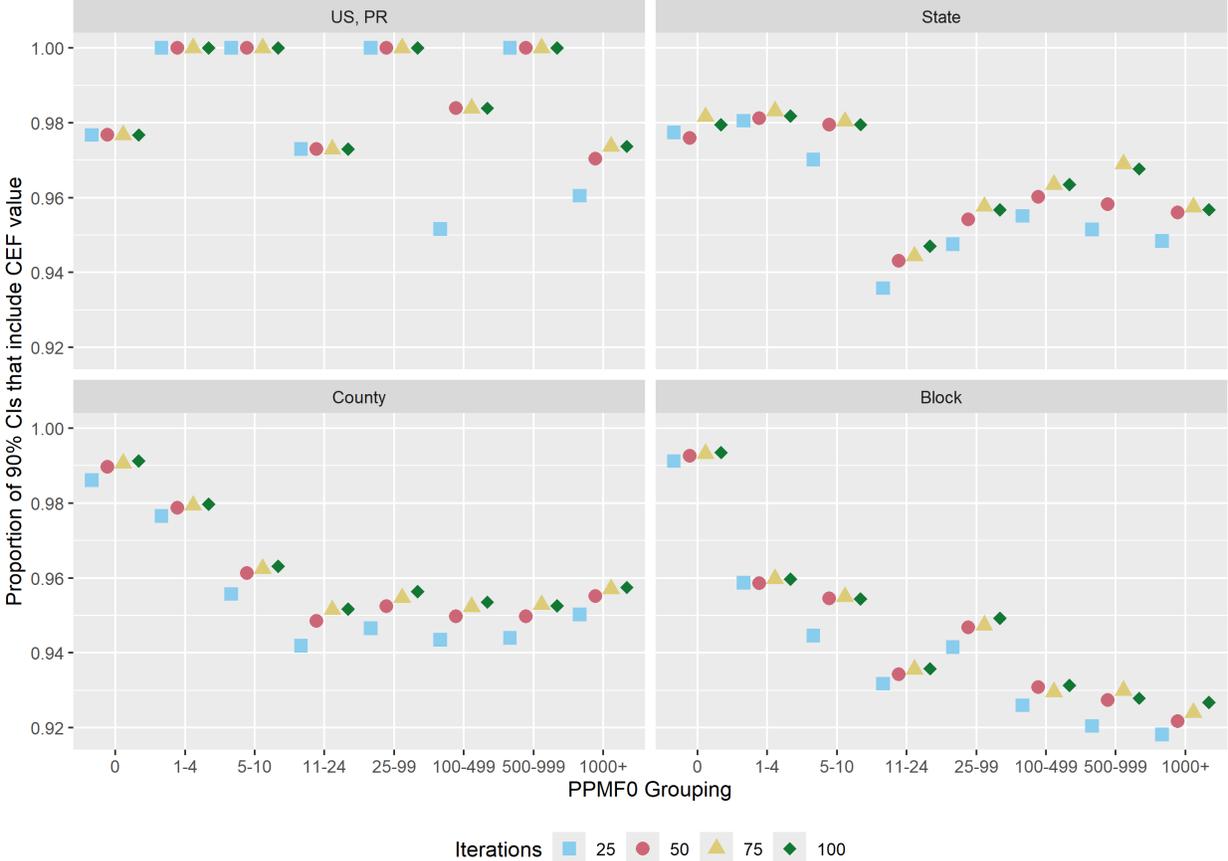
Figure 7: The proportion of 90% ct confidence intervals that contained the CEF value by geography, query size, and number of iterations. Color reflects the number of simulations in the analysis, that is, the different color points clustered together mean no significant dependence on the number of runs. The ordinate axis (Y) is the CI coverage and the abscissa axis (X) is the value of the query. The panels correspond to different geolevels. Reducing to 25 iterations decreased the proportion slightly in most of the groupings compared with using a higher number of iterations, but still generated intervals that met the 90% threshold.

4   Experimental results using 2010 Census Demographic and Housing Characteristics data

The Demographic and Housing Characteristics File (DHC) contains more granular tables than the redistricting tables. The DHC release is split into two data universes. The household universe (DHCH) provides counts of categories of householders and for categories of housing unit and group quarters types. The persons universe (DHCP) provides counts of individuals within population groups. Additional details on the TDA implemented for the DHC setting can be found in Cumings et al. [8]. In this set of experiments we considered 25 simulations of the TDA on the DHC tables using the $PPMF_0$ as the input.

The DHC file contains a very large set of tables. Instead of trying to consider all the tables, which would be too time-consuming, we selected a subset of tables for our experiments and considered all queries within those tables. The DHCP tables used were

- PCT8. Relationship by Age Under 18 Years

- PCT9, PCT9A-I. Household Type by Relationship for the Population 65 Years and Over

- PCT12. Sex by Single-Year Age

The tables for the race iterations correspond to A: White alone, B: Black or African American alone, C: American Indian and Alaska Native alone, D: Asian alone, E: Native Hawaiian and Other Pacific Islander alone, F: Some Other Race alone, G: Two or More Races, H: Hispanic or Latino, I: White alone, not Hispanic or Latino. Each DHCP table had multiple levels for a total of 405[8] unique queries per geography.

The DHCH tables used were

- H4. Tenure

- P19. Households by Presence of People 65 Years and Over, Household Size, and Household Type

- PCT7, PCT7A-I. Household Type by Household Size

- PCT10, PCT10A-I. Family Type by Presence and Age of Own Children

- PCT14, PCT14A-I. Presence of Multigenerational Households

Each DHCH table had multiple levels for a total of 405[9] unique queries per geography. Similar to the redistricting data experiments, we sampled geographic units by grouping by total population size and sampling 5 geographies per grouping per state in order to make the size of the results more manageable at the block and tract geographic levels.

4.1 DHCP results

Figure 8 shows the proportion of CIs that contained the CEF value by query size at the county level. Additional geographic levels are available in the online appendix. While overall we saw that the proportion was above the 90% threshold most of the time for multiple confidence interval types, the number of interval types that reached the threshold was lower relative to the redistricting tables. It was also more often the case that the intervals using the normal distribution quantile (z, BCz, cz) failed to meet the threshold. We observed more substantial estimated bias than the redistricting data experiments, so using bias-adjustments (and using the conditional rule) positively impacted the coverage more often. Among the query groups with higher count answers (500-999 and 1000+) at the U.S. and Puerto Rico level and block level, the bias-corrected confidence intervals were substantially better than the unadjusted intervals. The unadjusted intervals also missed the threshold in these cases. We theorize that this pattern is related to the privacy-loss budget allocation pattern used at these levels. Relatively small percentages of the privacy-loss budget (1.91% and 0.22%) were used at the U.S. and block levels, respectively for the DHCP [6]. As a result, those estimated queries would have larger variances and be more susceptible to bias related to enforcing non-negative counts. This pattern was also seen in the redistricting data, though it was less pronounced there. The redistricting data also used a relatively small portion of the budget at the U.S. and block levels (2.54% and 4.03%).

4.2 DHCH results

Among DHCH queries, we observed that the confidence intervals under-performed within some query size categories at the U.S. and Puerto Rico level but that pattern did not continue at lower geographies (Figure

---

[8]See 2020 Census Demographic and Housing Characteristics File Technical Documentation (available at: https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/ complete-tech-docs/demographic-and-housing-characteristics-file-and-demographic-profile/ 2020census-demographic-and-housing-characteristics-file-and-demographic-profile-techdoc.pdf for an explanation of the individual queries associated with each of the DHCP tables listed above.

[9]See 2020 Census Demographic and Housing Characteristics File Technical Documentation (available at: https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/ complete-tech-docs/demographic-and-housing-characteristics-file-and-demographic-profile/ 2020census-demographic-and-housing-characteristics-file-and-demographic-profile-techdoc.pdf for an explanation of the individual queries associated with each of the DHCH tables listed above.
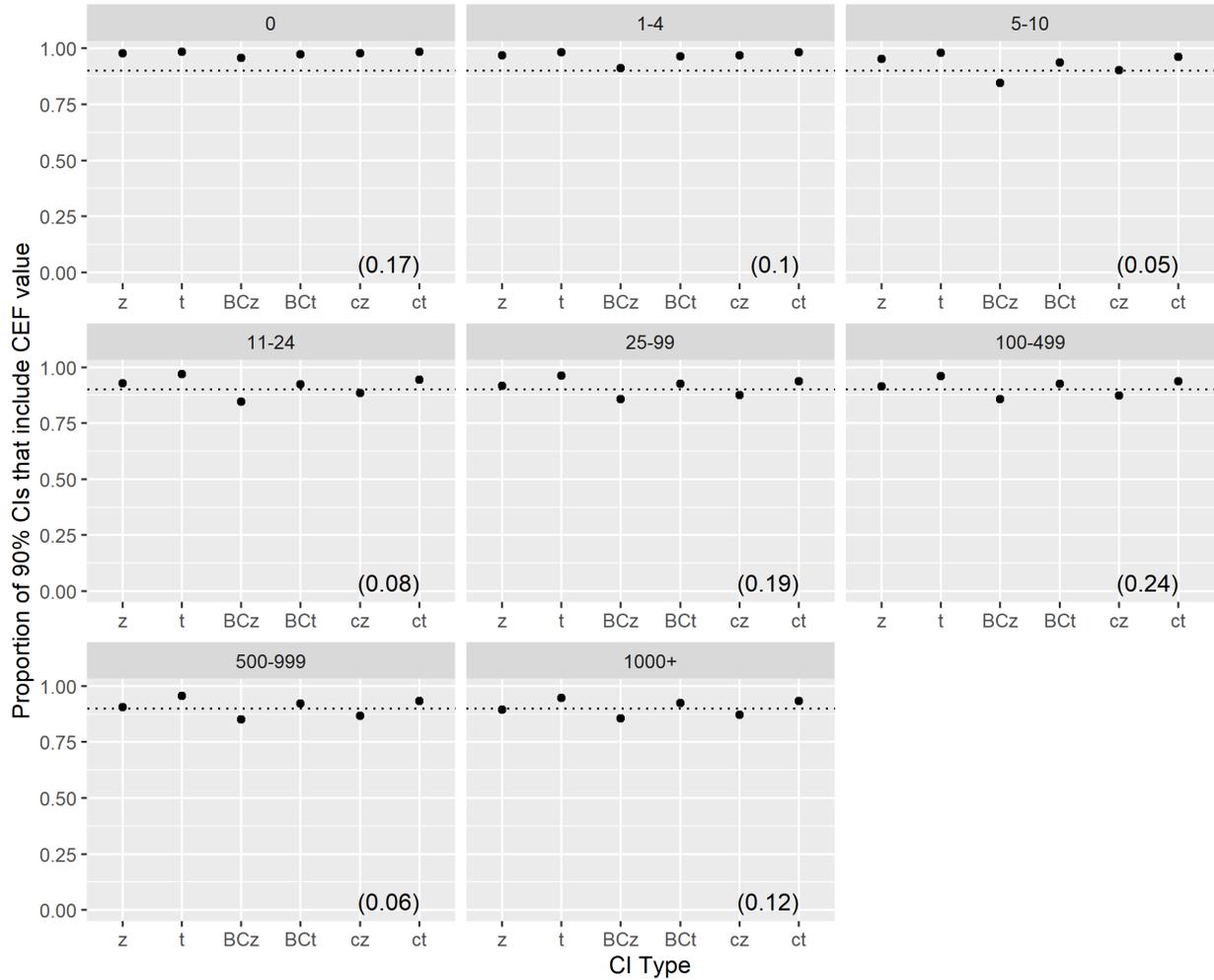
Figure 8: The Proportion of 90% CIs that contained the CEF value for 2010 County Level DHCP queries aggregated by query size (the bracketed number in the lower right corner indicates the proportion of the queries in that size category). Mostly, the proportion is above 0.9 threshold, but generally the proportions are lower than the redistricting queries; also CIs using normal distribution (z, BCz, cz) fail to reach the threshold more often.

9). For example, county level queries (Figure 10) were either very close (z, BCz, cz) or exceeded (t, BCt, ct) the 90% threshold across all the query size categories. At the U.S. Level, most of the queries (81%) were in the largest size category (1000+) where the bias-corrected or conditionally bias-corrected intervals met the 90% threshold. So overall, queries for which the confidence intervals under-performed represented a small proportion of all queries. We hypothesize that the reason we see the worse performance is the relatively small proportion of privacy-loss budget spent at the U.S. level (6.28%) [6], which could result in biased TDA query estimates because of the effect of the the non-negative count constraint. Relatedly, at the U.S. and Puerto Rico level, the bias correction for the confidence intervals was beneficial for all query size categories 25 and above and made a substantial improvement for the largest queries.

4.3   Estimated bias

In order to better understand the link between the confidence intervals and the bias, we computed a summary of the estimated bias using the approximate Monte Carlo method (bias relative to the $PPMF_0$ answers).
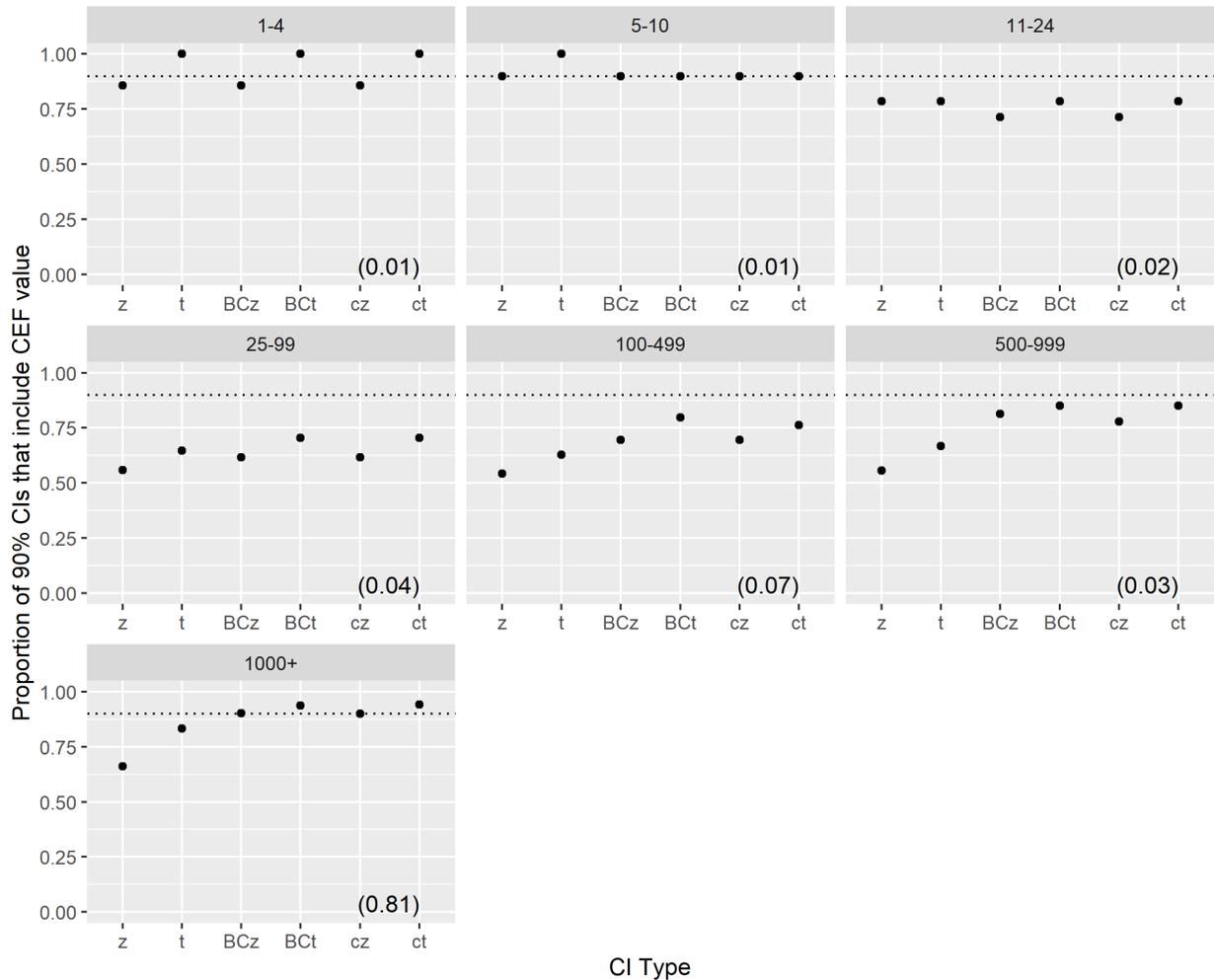
Figure 9: The proportion of 90% CIs that contained the CEF value for 2010 U.S. and Puerto Rico Level DHCH queries aggregated by query size (the bracketed number in the lower right corner indicates the proportion of the queries in that size category, e.g. in this figure most – 0.81 – have values over 1000). US level queries often underperform, likely due to lower privacy budget allocation; bias correction is beneficial for high count query groups (500+).

Tables 2 and 3 show the 1st, 50th, and 99th percentiles of the estimated bias by geographic level and query size group for DHCH queries (similar tables for DHCP are available in the online appendix). We see that the 1st and 99th percentiles of estimated bias at the US level for the 1000+ grouping are substantially more extreme than those at the state level.

The estimated bias also effects the width of the confidence intervals because the RMSE is used in the margin of error calculation. As the estimated bias of a query increases, so does the width of the confidence interval. We know from Section 3.1 that the bias can be underestimated by the AMC methodology, so even when we are using bias-adjustments to the confidence intervals, allowing for a wider margin of error is beneficial in that it will help to make up for underestimation. However, there may be cases in which the magnitude of the estimated bias and its effect on the width of the resulting confidence intervals is so large that the confidence interval is no longer meaningful. For example, the AMC-estimated bias for the U.S. level query concerning the number of households with no one over 65 years of age, as provided in Table P19, is $-41,137$ with a BCt 90% CI of $(87,517,612 - 87,683,422)$. In these cases, it may be preferable to utilize the noisy measurement files to get an unbiased estimate of the query and a confidence interval derived from the
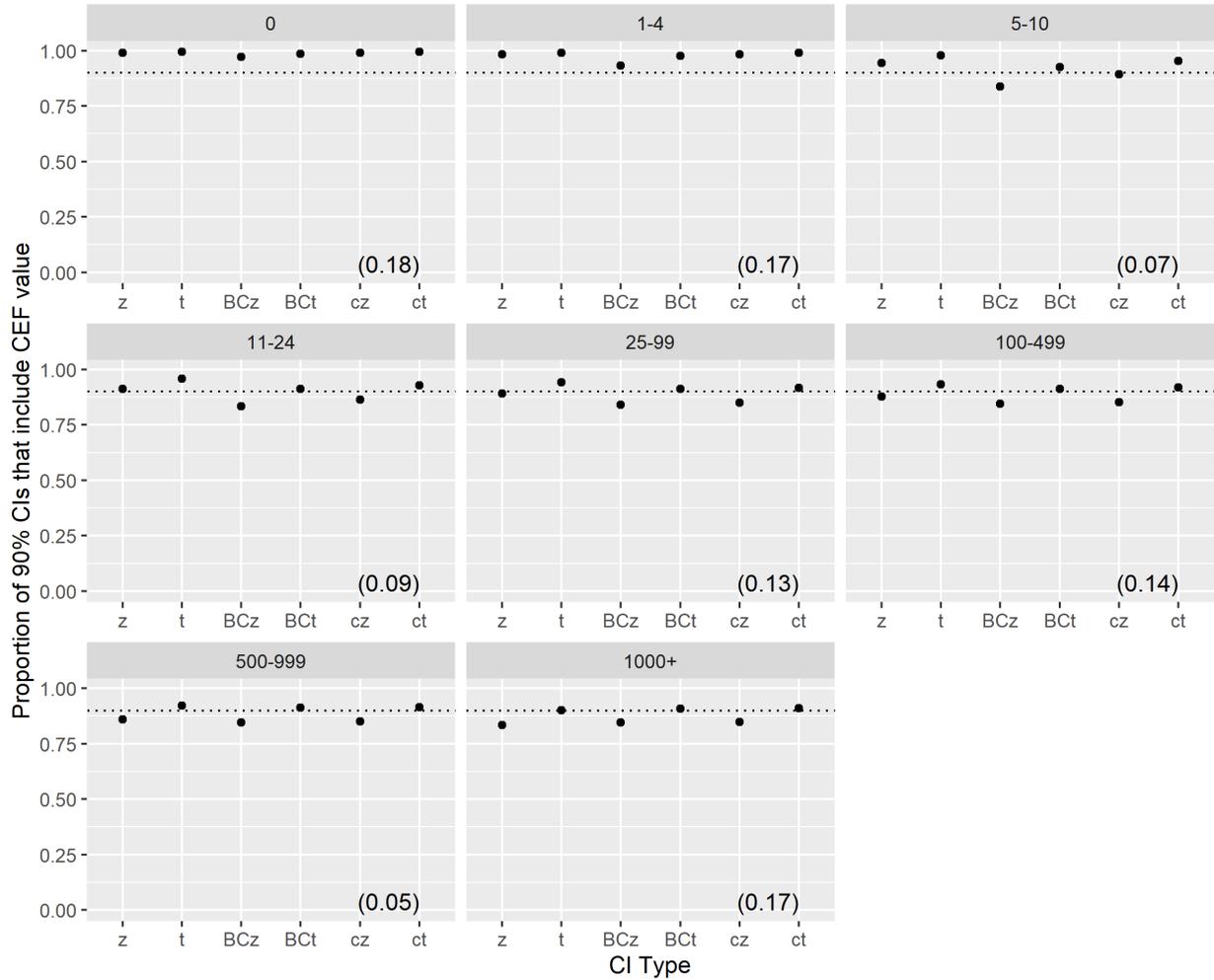
Figure 10: The proportion of 90% CIs that contained the CEF value for 2010 county level DHCH queries aggregated by query size (the bracketed number in the lower right corner indicates the proportion of the queries in that size (count) category. The proportion of CIs containing the CEF value is above the 0.9 threshold for the t, BCt, and ct intervals

| | US | | | State | | | County | | | Tract | | |
Query Size	1st	50th	99th	1st	50th	99th	1st	50th	99th	1st	50th	99th
0				0.4	1.3	3.9	0.0	0.3	3.0	0.0	0.2	2.2
1-4	-0.4	2.6	6.4	-1.3	1.2	6.1	-2.2	0.1	4.6	-2.3	-0.2	3.5
5-10	-2.9	4.3	6.7	-3.0	1.6	9.0	-3.9	0.0	6.2	-3.9	-0.3	4.8
11-24	-4.7	6.1	20.6	-4.8	1.6	15.4	-5.8	0.1	8.1	-5.6	-0.2	6.1
25-99	-4.9	12.0	45.6	-7.5	3.4	26.1	-8.8	0.0	11.5	-8.2	-0.3	8.1
100-499	-20.0	10.6	126.4	-16.2	3.3	49.8	-14.6	-0.4	15.7	-12.2	-0.2	12.1
500-999	-30.7	-0.9	225.1	-28.4	1.6	52.9	-21.2	-0.8	21.3	-14.5	-0.1	12.9
1000+	-6321.2	-4.9	4460.1	-636.6	-1.6	510.1	-62.2	-0.5	55.2	-16.1	-0.1	11.7

Table 2: The 1st, 50th and 99th percentile estimated bias of DHCH queries by query size for US, state, county, and tract geographies

| | Block | | | AIAN Area | | | School District | | |
Query Size	1st	50th	99th	1st	50th	99th	1st	50th	99th
0	0.0	0.0	2.0	0.0	0.1	2.0	0.0	0.1	2.2
1-4	-3.0	-0.6	5.5	-2.2	-0.2	4.1	-2.3	-0.2	4.2
5-10	-6.1	-1.3	8.0	-3.9	-0.2	6.1	-4.1	-0.2	6.2
11-24	-10.0	-1.4	10.2	-6.1	-0.4	8.7	-6.4	-0.2	8.3
25-99	-14.3	-0.7	15.4	-10.8	-0.5	13.4	-10.5	-0.2	12.1
100-499	-17.7	1.7	23.0	-21.6	-0.6	26.2	-18.1	-0.4	19.1
500-999	-10.7	2.9	24.1	-35.1	-1.4	45.5	-27.3	-0.5	28.2
1000+	-9.9	0.3	4.7	-97.2	-1.7	88.6	-102.0	-0.5	76.3

Table 3: The 1st, 50th and 99th percentile estimated bias of DHCH queries by query size for Block, AIAN area, and elementary school district geographies

noisy measurement(s). Estimates and confidence intervals calculated directly from the noisy measurements will not benefit from the error reduction that TDA's post-processing affords. That said, for estimates at higher geographic levels (especially at the U.S. and state level where the benefit of TDA's error reduction is less pronounced) for queries derived from noisy measurements with larger shares of privacy-loss budget, we would anticipate the confidence intervals derived from the noisy measurements to outperform those generated via the AMC approach.

5   Discussion

We proposed a general method for estimating key statistical properties of query estimates from the TDA and how to combine them into meaningful confidence intervals. While we cannot examine the results for every single possible query at every single level of geography, the results consistently demonstrate that the AMC method performed well using the 2010 data and a range of queries using as few as 25 simulations replicates. Given that the 2010 parameters mirrored those used for the 2020 Census, we anticipate this methodology will work similarly well using the 2020 Census data.

Among the proposed confidence interval types, we found that Wald-type intervals performed better than the non-parametric quantile intervals. Within the Wald-type intervals, we found that bias-adjusted intervals were generally less effective overall, and only select cases yielded improved coverage results. We attempted to characterize the select instances in which the bias-adjusted estimator may improve coverage with a conditional rule.

Considering the results overall, we recommend users utilize the conditional bias-corrected intervals with the Student's t distribution. Future work might improve the logic of that rule, but we think the experiments show that it does a good job balancing when to use or not use the bias correction. Using the more conservative Student's t critical value was often necessary to reach the desired threshold for DHC data table queries, and therefore we recommend its use over the normal distribution critical value.

A major benefit of this methodology is that it does not incur any additional privacy loss because it does not access confidential data. The method is also relatively simple to apply for a single query or set of queries, assuming the user can compute or retrieve the query value computed for the $PPMF_0$ and the associated simulations. Since all the inputs of our proposed AMC simulation approach (*i.e.*, the PPMF, the TopDown code, and parameters) are public[10] is also feasible for an organization outside of the Census Bureau to create additional simulation runs in the event this would be beneficial for a particular use case. There are a wide variety of applications or extensions where a version of the proposed methodology could be successfully deployed. We have have focused here on count queries, but for example ratio queries (e.g. persons per household) may be of interest. In addition, more complex statistical quantities such as the covariance

---

[10]The 2010 $PPMF_0$ and associated simulations are available at `https://registry.opendata.aws/census-2010-amc-mdf-replicates/`. Additional resources including a Jupyter notebook containing a full working code example for estimating the bias-corrected intervals with the Student's t distribution are available via the U.S. Census Bureau at `https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/demographic-and-housing-characteristics-file-and-demographic-profile/data_analysis_resources/1_estimating_conf_intervals_2010_demo/` and `https://github.com/uscensusbureau/AMC_Confidence_Intervals`.

between two different queries could be estimated using the multiple iterations, or even the uncertainty of a query total could be taken into account in a linear regression through a strategy similar to multiple imputation combination rules that incorporates the variation between and within different replicates [10]. Future research is needed to explore a more appropriate way of constructing confidence intervals when the distribution of the query answers across the iterations is not approximately normally distributed (such as when the query count is near 0 or if ratios were considered). While we focused specifically on the feasibility of this methodology for the TDA and the 2020 Census use case, the Approximate Monte Carlo method could apply more broadly to any algorithm whose output produces data that could then be used as the input to the algorithm.

References

[1] John Abowd et al. *2020 Census Demographic and Housing Characteristics (DHC) Noisy Measurement File.* 2023. URL: `https://www2.census.gov/programs-surveys/decennial/2020/data/demographic-and-housing-characteristics-file/00-2020-DHC-Noisy-Measurement-File/Noisy-Measurement-File-2020-Census-DHC.pdf`.

[2] John Abowd et al. *2020 Census Redistricting Data (P.L. 94-171) Noisy Measurement File.* 2023. URL: `https://www2.census.gov/programs-surveys/decennial/2020/data/01-Redistricting_File--PL_94-171/00-2020-Redistricting-Noisy-Measurement-File/Noisy-Measurement-File-2020-Census-Redistricting-Data.pdf`.

[3] John Abowd et al. "The 2020 Census Disclosure Avoidance System TopDown Algorithm". In: *Harvard Data Science Review* Special Issue 2 (June 2022). https://hdsr.mitpress.mit.edu/pub/7evz361i.

[4] Robert Ashmead. *Estimating the Variance of Complex Differentially Private Algorithms.* Joint Statistical Meetings. 2019. URL: `https://www.census.gov/content/dam/Census/newsroom/press-kits/2019/jsm/presentation-estimating-the-variance-of-complex-differentially-private-algorithms.pdf`.

[5] Thomas Brawner and James Honaker. "Bootstrap inference and differential privacy: Standard errors for free". In: *Unpublished Manuscript* (2018).

[6] U.S. Census Bureau. *Demonstration Data Products Suite 2023-04-03 Privacy Loss Budget Allocations.* 2023. URL: `https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/04-Demonstration_Data_Products_Suite/2023-04-03/2023-04-03_Privacy-Loss_Budget_Allocations.pdf`.

[7] Karan Chadha, John Duchi, and Rohit Kuditipudi. "Resampling methods for Private Statistical Inference". In: *arXiv preprint arXiv:2402.07131* (2024).

[8] Ryan Cumings-Menon et al. *Disclosure Avoidance for the 2020 Census Demographic and Housing Characteristics File.* 2023. arXiv: `2312.10863 [cs.CR]`.

[9] Cecilia Ferrando, Shufan Wang, and Daniel Sheldon. "Parametric bootstrap for differentially private confidence intervals". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2022, pp. 1598–1618.

[10] D. B. Rubin. *Multiple Imputation for Nonresponse in Surveyes.* New York: Wiley, 1987.

[11] Zhanyu Wang and Jordan Awan. "Debiased Parametric Bootstrap Inference on Privatized Data". In: URL: `https://api.semanticscholar.org/CorpusID:265313051`.

[12] Zhanyu Wang, Guang Cheng, and Jordan Awan. "Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies". In: *arXiv preprint arXiv:2210.06140* (2022).

6   Appendix: Examples of confidence intervals

In order to be illustrative of what the associated confidence intervals for specific queries look like, we give examples of the conditionally bias-corrected confidence intervals for specific queries from the experiments,

6.1 Redistricting queries

Note that the following redistricting queries are constructed from 25 AMC iterations rather than the 100 used in the experiments.

QUERY	State	PPMF Value	Est. Bias	Est. RMSE	90% CI
P0010065	05	4	-0.68	3.45	(0 - 11)
P0020016	20	831	2.96	6.85	(817 - 845)
P0020036	33	121	-2.56	5.78	(109 - 133)
P0030004	02	16895	-1.44	13.72	(16,867 - 16,923)
P0030013	32	13574	-0.92	7.49	(13,558 - 13,590)
P0030062	49	1	1.12	2.38	(0 - 6)
P0040010	20	1603	2.40	6.40	(1,590 - 1,616)
P0040032	44	163	-2.08	4.76	(153 - 173)
P0040036	49	714	-1.84	7.23	(699 - 729)
P0040043	51	61	0.00	5.42	(50 - 72)

Table 4: Example 90% confidence Intervals for 2010 State Level Redistricting Queries

QUERY	State	County	Tract	PPMF Value	Est. Bias	Est. RMSE	90% CI
H0010001	40	109	103400	167	0.00	0.00	(167 - 167)
P0010010	18	141	011305	107	-2.20	6.26	(94 - 120)
P0010012	38	089	964000	7	-0.32	2.04	(2 - 12)
P0010028	12	011	100300	4	-1.32	2.37	(0 - 9)
P0010063	53	035	090202	4	-2.88	3.02	(0 - 11)
P0020003	21	111	002700	2923	4.04	6.19	(2,906 - 2,932)
P0020016	49	035	111802	12	0.00	2.24	(7 - 17)
P0020029	42	003	415001	2	-0.80	1.55	(0 - 6)
P0030010	13	285	960502	33	-0.64	4.80	(23 - 43)
P0030013	41	051	004900	35	-0.24	2.91	(29 - 41)

Table 5: Example 90% Confidence Intervals for 2010 Tract Level Redistricting Queries