1 Sensitivity analysis for verification measures

The performance of a differentially private verification measure is sensitive to variations in the data structure and in the values of the parameters of the method that generates said verification measure. In this section, we conduct a sensitivity analysis of the ϵ-differentially private verification measure for regression coefficients described in subsection 4.1. of the main text. We consider how different data structures and values of the parameters in the method that generates ϵ-differentially private verification measure influence its performance.

We begin this section by describing the parameters involved in the method to generate ϵ-differentially private verification measure. Next, we explain how different data structures and values of these parameters can influence the performance of the verification measure. Since certain data structures can make some regression coefficients nonestimable, we provide and describe a new verification measure that reports the number of nonestimable coefficients and the number of coefficients exceeding the differential private threshold. Finally, we provide a sensitivity analysis of the performance of both the ϵ-differentially private verification measure and the new verification measure

In Subsection 4.1 of the main text, we provide an ϵ-differentially private verification measure for regression coefficients. The measure allows analysts to assess whether or not the value of a regression coefficient β_j exceeds some threshold, γ_0. To achieve ϵ-differential privacy, we use the subsample and aggregate method proposed by Nissim et al. (2007). This method requires spliting the data $\mathbf{D}$ into M subsets. Analysts then have to select values for ϵ and M to compute the proposed measure. Different values of ϵ and M can impact the measure's performance in distinct ways. At the same time, the effect of values of ϵ and M on the measure's performance interacts with other factors such as the sample size of $\mathbf{D}$—denoted by n—and n_j, where n_j is the number of observations in $\mathbf{D}$ available to estimate β_j. The role of n_j is relevant when β_j relates to a level of a categorical variable. For example, if β_j is associated with American Indian/Alaska Native race, n_j would equal the number of employees that chose to identify with that race. Thus, we need to consider how ϵ, M, n, and n_j influence the performance of the ϵ-differentially private verification measure for regression coefficients.

Distinct combinations of different values of ϵ, M, n, and n_j lead to divergent outcomes. Large values of ϵ increase privacy levels, but decrease the statistical usefulness of the verification measures. M controls the sensitivity of the verification measure: the larger the M, the lower the sensitivity and vice-versa. Verification measures with low sensitivity require less ϵ to retain suitable usefulness. However, if we randomly split $\mathbf{D}$ into too many subsets, $\mathbf{D}_1, \ldots, \mathbf{D}_M$, the standard error of the estimates of β_j—within subsets— increases. When $N = M/n$ is small, the increase in the estimates standard error comes to the detriment of their usefulness. Hence, the definition of what we can consider as a large M also depends on n.

n_j is relevant when β_j represents a level of a categorical variable. For example, if β_j is associated with American Indian/Alaska Native race, n_j would equal the number of employees that chose to identify with that race. If β_j represents a level of a categorical variable and M is large, small values of n_j can lead to high standard errors. In addition, we can end up not having observations to estimate β_j within some subsets. That is, we can have some $\mathbf{D}_l$ with $n_{lj} = 0$, where n_{lj} is the number of observations in $\mathbf{D}_l$ available to estimate β_j. Another situation where β_j will be nonestimable is when $\mathbf{D}_l$ does not have information to estimate the intercept of the regression model. This could happen if $\mathbf{D}_l$ does not contain observations for at least one of the

reference levels of the categorical variables. When β_j is nonestimable from $\mathbf{D}_l$, the software will produce an error when computing b_{jl}, where b_{jl} is the MLE estimate of β_j computed from $\mathbf{D}_l$.

Motivated by the nonestimability of β_j for some $\mathbf{D}_l$, we provide a new verification measure for β_j. The new measure allows us to report, in a differential private way, the number of b_{jl}s that we were not able to estimate and the number of b_{jl} that exceed the differential private threshold. In what follows, we describe the new verification measure for β_j. Then, we provide a sensitivity analysis of the performance of both ϵ-differentially private verification measure and the new verification measure for β_j.

1.1 Measures for importance of estimable and nonestimable regression coefficients

Similarly to Subsection 4.1 of the main text, we aim to infer whether or not $\theta_0 = \mathbb{I}_{(-\infty,\gamma_0]}(\beta_j)$ is equal to zero or one, where β_j is a coefficient associated with a level of a categorical variable. To do so, we plan to approximate $\theta_0 = \mathbb{I}_{(-\infty,\gamma_0]}(\beta_j)$ by using the pseudo-parameter,

$$\theta_N = \begin{cases} 1 & \text{if } P(\hat{\beta}_j^N \leq \gamma_0 | \hat{\beta}_j^N \neq \texttt{NA}\ ) \geq \gamma_1, \\ 0 & \text{if } P(\hat{\beta}_j^N \leq \gamma_0 | \hat{\beta}_j^N \neq \texttt{NA}) < \gamma_1, \end{cases}$$

where $\hat{\beta}_j^N$ is the MLE of β_j based on a sample with N individuals (where N stands for a generic sample size). Here, $\gamma_1 \in (0,1)$ reflects the degree of certainty required by the user before she decides there is enough evidence to conclude that $\theta_0 = 1$. We also assume that $\hat{\beta}_j^N \in \mathbb{R} \cup \{\texttt{NA}\}$, where NA represents the nonestimability condition. Recall that $\hat{\beta}_j^N = \texttt{NA}$ if, for example, $n_j = 0$. When $\hat{\beta}_j^N$ is a consistent estimator of β_j, we can guarantee that $\lim_{N\to\infty} \theta_N = \theta_0$ if $\lim_{N\to\infty} P(\hat{\beta}_j^N \neq \texttt{NA}) = 1$.

Unfortunately, we cannot release $\hat{\beta}_j^N$, nor other deterministic functions of $\mathbf{D}$, directly. Instead, we release a noisy version of the key quantity in θ_N, namely $\mathbf{q} = (q_1, q_0, q_{\texttt{NA}}) = (P(\hat{\beta}_j^N \leq \gamma_0), P(\hat{\beta}_j^N > \gamma_0), P(\hat{\beta}_j^N = \texttt{NA}))$. We do so using the sub-sample and aggregate method (Nissim et al., 2007). We randomly split $\mathbf{D}$ into M disjoint subsets, $\mathbf{D}_1, \ldots, \mathbf{D}_M$, of size N (with inconsequential differences when $N = n/M$ is not an integer), where M is selected by the user. In each $\mathbf{D_l}$, where $l = 1, \ldots, M$, we compute the MLE b_{jl} of β_j. The $(b_{j1}, \ldots, b_{jM})$ can be treated as M independent draws from the distribution of $\hat{\beta}_j^N$, where $N = n/M$. Let $\mathbf{W}_l = (\mathbb{I}_{(-\infty,\gamma_0]}(b_{jl}),\ \mathbb{I}_{(\gamma_0,\infty)}(b_{jl}),\ \mathbb{I}_{\texttt{NA}}(b_{jl}))$. Each W_l is an independent, multinomial distributed random variable with parameters one and $\mathbf{q}$. Thus, inferences for $\mathbf{q}$ can be made based on $\mathbf{S} = \sum_{l=1}^{M} \mathbf{W}_l$. However, we cannot release $\mathbf{S}$ directly and satisfy ϵ-DP; instead, we generate a noisy version of $\mathbf{S}$ using the Laplace Mechanism with $\lambda = 2/\epsilon$, resulting in $\mathbf{S}^R = \mathbf{S} + \boldsymbol{\eta}$. The global sensitivity equals 2, since at most two components of $\mathbf{S}$ can change by at most one unit .

The noisy $\mathbf{S}^R$ satisfies ϵ-DP; however, interpreting it directly can be tricky. First, the components of $\mathbf{S}^R$ are not guaranteed to lie in $(0, M)$ nor even to be an integer and to add to M. Second, alone $\mathbf{S}^R$ does not provide estimates of uncertainty about $\mathbf{q}$. We therefore use a post-processing step—which has no bearing on the privacy properties of component of $\mathbf{S}^R$—to improve interpretation. We find the posterior distribution of $\mathbf{q}$ conditional on $\mathbf{S}^R$ and using the noise distribution, which is publicly known. Using simple MCMC techniques, we estimate the

model,

$$\mathbf{S}^R|\mathbf{S} \sim \text{Laplace}_3(\mathbf{S}, 2/\epsilon), \quad \mathbf{S} \mid \mathbf{q} \sim \text{Binomial}(M, \mathbf{1}), \quad \mathbf{q} \sim \text{Dirichlet}(1, 1, 1), \tag{1}$$

where Laplace_3 denote the multivariate distribution induced by three independent univariate Laplace distributions with location and scale parameters determined by $\mathbf{S}$ and $2/\epsilon$. Here, we treat $\mathbf{S}$ as an unobserved random vector and average over it.

The verification server reports back the posterior distribution of $\mathbf{q}$ to the analyst, who can approximate θ_N for any specified γ_1 simply by finding the amount of posterior mass below γ_1. Alternatively, analysts can interpret, for example, the posterior distribution for $q_1/(q_1 + q_0)$ as a crude approximation to the Bayesian posterior probability, $\pi\left(\beta_j \leq \gamma_0 | \mathbf{S}^R\right)$. For instance, if the posterior mode for $q_1/(q_1 + q_0)$ equals 0.87, we could say that the posterior probability that $\beta_j < \gamma_0$ is approximately equal to 0.87. Finally, the analyst can use $q_{\texttt{NA}}$ to quantify the estimability of $\hat{\beta}_j$ among the subsets $\mathbf{D}_1, \ldots, \mathbf{D}_M$. We caution that the analyst should refrain from providing conclusions based on $q_1/(q_1 + q_0)$ on those cases where $q_{\texttt{NA}}$ is large.

1.2 Sensitivity study for verification measures on race wage gaps

We present a sensitivity analysis for the verification measures proposed in Subsection 1.1 and Subsection 4.1 of the main text; hereafter, we refer to these measures as multinomial- and binomial-based measures, respectively. We perform a sensitivity analysis in order to: (1) determine how different values of ϵ, M, n, and n_j impact the verification measures' performance, (2) compare the accuracy of the two proposed verification measures, and (3) identify and design strategies that guide analysts better use and interpret the verification measures.

As mentioned before, our sensitivity study aims to assess how different values of ϵ, M, n, and n_j impact the verification measures' performance. To do this, we measure the sensitivity of the multinomial- and binomial-based verification measures across seven scenarios. The scenarios are defined by subsetting the Status File and determining which regression coefficient will be verified. Each subset of the Status File leads to distinct values of n and n_j. Then, for each subset we perform a regression analysis to obtain regression coefficients. Next, we query the multinomial- and binomial-based measures using $\epsilon \in \{.5, 1, 2\}$ and $M \in \{10, 30, 50\}$. To account for the variability derived from the random mechanisms that partition the datasets and generate the Laplace error, for each scenario and combination of ϵ and M, we query ten times the verification measures. Thus, we can report average values and standard errors over queries to assess the verification measures performance.

In what follows, we describe in more detail our strategy to assess the verification measures performance. Then, we describe the results of our sensitivity study. Finally, we provide some general conclusions and recommendations for using and interpreting the verification measures.

This sensitivity study considers seven scenarios displayed in Table 1. We assemble each scenario by subsetting the Status File according to different combinations of agency, occupation, and sex that lead to distinct values of n and n_j. For each subset, we perform an overall regression analysis similar to that described in Subsection 5.3 of the main text where each observation is an employee-year. The dependent variable is the natural logarithm of each employee's inflation adjusted basic pay in a given year. The key independent variable is the race with which individual employees identify. While we include other independent variables such as employees' age as well

as its square, and years of education, race is the specific coefficient to be verified from the overall analysis.

The examined scenarios consider three different sample sizes: large, if $n > 100,000$; moderate, if $1,000 < n \leq 100,000$; and small if $n \leq 1,000$. Regarding the number of employees whose race will be verified, we also consider three different sizes: large, if $n_j > 10,000$; moderate, if $1,000 < n_j \leq 10,000$; and small if $n_j \leq 1,000$. For our sensitivity study, we assume that the aim is to verify whether or not the coefficient is below -0.01. We choose scenarios considering three different situations: coefficient estimate meets this condition, coefficient estimate does not meet this condition, and estimate is approximately equal to the threshold. For each scenario, we query the multinomial- and binomial-based measures using $\epsilon \in \{.5, 1, 2\}$ and $M \in \{10, 30, 50\}$. To account for the variability derived from the random mechanisms that partition the datasets and generate the Laplace error, for each scenario and combination of ϵ and M, we query ten times the verification measures. We focus on the average values and standard errors over queries to assess the verification measures performance.

Scen	Agency	Occupation	Sex	Race	n	n_j	$\hat{\beta}_j$
1	VATA	Practical Nurse	F	Black	193,144	66,394	0.02
2	SZ00	Social Insurance Administration	F	Hispanic	305,311	42,709	-0.04
3	TD03	Air Traffic Control	M	Black	458,114	21,416	-0.05
4	TR93	Tax Examining	F	Asian	165,210	3,963	-0.03
5	DJ09	General Attorney	F	Black	34,154	3,748	-0.01
6	DJ02	Criminal Investigating	F	Asian	42,077	1,225	-0.04
7	LF00	General Attorney	F	Asian	705	32	-0.05

Table 1: Scenarios' (Scen) definition. Agency, occupation, and sex define the subset of the Status File to be analyzed. Race determines the coefficient to be verified and $\hat{\beta}_j$ its corresponding MLE estimate. The considered agencies are Veterans Health Administration (VATA), Social Security Administration (SZ00), Federal Aviation Administration (TD03), Internal Revenue Service (TR93), Office of U.S. Attorney (DJ09), Federal Bureau of Investigation (DJ02), and Federal Election Commission (LF00). n denotes the number of employees in each subset and n_j indicates the number of employees identifying with the displayed race (fifth column). $\hat{\beta}_j$ provides the coefficient estimates from overall regression models.

First, we examine the average values of the coefficients obtained after ten queries. We base our analysis on $\hat{r}$ and $\hat{q}$, where $\hat{r}$ and $\hat{q}$ are the posterior modes of r and $q_1/(q_1+q_0)$, respectively. Table 2 displays the average values—over ten queries—of $\hat{r}$ and $\hat{q}$ for all scenarios and values of ϵ and M. We start by analyzing the results for Scenarios 5 and 7 because they represent special situations with very small values for n_j and coefficients that are not unambiguously above or below the threshold. Then, we present results for the remaining scenarios.

In Scenario 5, the coefficient estimate is approximately equal to the threshold. Hence, we expect that $\hat{r}$ and $\hat{q}$ are around .5 as reflected in Table 2. For this scenario, there are no obvious patterns that we can identify for the average values of $\hat{r}$ and $\hat{q}$ as M and ϵ changes. In Scenario 7, the number of employees who identify as Asian is very small. Therefore, the subsample and aggregate method often leads to subsets having none employees who identify as Asian. The coefficient for race is less than -.01 (below the threshold) under Scenario 7. We would expect that a coefficient below -.01 results in large values of $\hat{r}$. However, Table 2 shows small values of $\hat{r}$

if we use the binomial-based measure. If we use the multinomial-based measure, $\hat{q}$ is around .5. Thus, it is not evident whether or not we could claim that the coefficient is below the threshold.

Scenarios 1, 2, 3, 4, and 6 grant manageable values for n_j and coefficients unambiguously above or below the threshold. For these scenarios, we observe that the average values of $\hat{r}$ and $\hat{q}$ tend to provide an accurate assessment of the tested condition. Nonetheless, the results obtained under different scenarios allow us to shed light on how different values of n, n_j, ϵ, and M impact the verification measures' performance. We observe that when the sample size n is large—such as under Scenario 1—and the proportion of employees who identify with the race of interest n_j/n is also large—34% in Scenario 1—we obtain more accurate results as M increases regardless the value of ϵ. However, if n_j/n is small—$\leq$5%—such as in Scenarios 3, 4, 6, and, for $\epsilon = 1,\ 2$, the verification measures become less accurate as M increases. There are other cases—e.g., Scenario 2 with $\epsilon = 1,\ 2$—where n and n_j/n are large and results are more accurate when $M = 30$. For these cases, we conjecture that $M = 50$ increases the standard error of the estimates negatively affecting the performance of the verification measures.

We can also notice in Table 2 that, in many cases, measures behave—as a function of M—differently when $\epsilon = .5$ and $\epsilon = 1,\ 2$, and no standard pattern is easy to identify. This is not surprising since to set $\epsilon = .5$ leads to inject more noise to the verification measures. Accuracy also depends on whether we use the binomial or multinomial-based strategy. Unfortunately, we cannot conclude that one of the considered measures outperforms the other. We can only observe that the absolute difference between average values of $\hat{r}$ and $\hat{q}$ decreases as ϵ increases—maximum differences are .07, .05, .03 when ϵ is .5, 1, and 2, respectively. Since the observed differences are small, we can conclude that, for scenarios like 1, 2, 3, 4, and 6, the average performance of $\hat{r}$ and $\hat{q}$ is indistinctly the same.

Until now, we have focused on examining the average values of the coefficients. We can also assess the standard errors of the coefficients to assess the verification measures performance. Table 2 also displays standard errors—over the ten queries—for $\hat{r}$ and $\hat{q}$. As expected, we observe that most of the times standard errors reduce or stay similar when ϵ increases. We also observe a similar pattern when M increases. Recall that, if M increases, there is an underlying trade-off between the variance of the Laplace mechanism and the variance of the estimator of β_j. Hence, it seems that when M increases, the rate at which the variance of the Laplace mechanism decreases is faster than the rate at which the variance of the estimator of β_j increases within partitions. On the other hand, for $\epsilon = 0.5$, several standard errors are fairly large. For $\epsilon = 1, 2$, most of the times standard errors are small except when $M = 10$—standard errors up to .232. Finally, we also notice that most of the times the standard errors for the multinomial-based measure are higher than those for the binomial-based measure. This is expected since with the multinomial-based measure releases more information, meaning that more error needs to be added to the outcome.

The problem of having nonestimable regression coefficients is relevant and requires special attention. The multinomial-based strategy provides a mean to retrieve, using differential privacy, the proportion of nonestimable regression coefficients within a partition of size M. Recall from Subsection 1.1 that we denote this proportion by $q_{\texttt{NA}}$. Table 3 displays the average values and standard error—over ten queries—of the posterior modes of $q_{\texttt{NA}}$ —denoted by $\hat{q}_{\texttt{NA}}$— for all scenarios and values of ϵ and M. For the first six scenarios, we observe that as either M or ϵ increases, the average $\hat{q}_{\texttt{NA}}$ decreases with few exceptions when epsilon is .5. The magnitude of $\hat{q}_{\texttt{NA}}$ for these scenarios is always small. For Scenario 7, as expected, we notice that as M increases

				$\epsilon = .5$		$\epsilon = 1$		$\epsilon = 2$	
E	$\hat{\beta}_j$	S_1	M	$\hat{r}$	$\hat{q}$	$\hat{r}$	$\hat{q}$	$\hat{r}$	$\hat{q}$
		0	10	.07 (.066)	.09 (.070)	.04 (.061)	.11 (.193)	.04 (.063)	.02 (.022)
1	.02	0	30	.03 (.051)	.04 (.052)	.01 (.016)	.04 (.044)	.01 (.014)	.01 (.003)
		.3	50	.03 (.035)	.02 (.016)	.02 (.027)	.01 (.014)	.01 (.017)	.01 (.008)
		10	10	.94 (.088)	.84 (.155)	.93 (.073)	.92 (.071)	.98 (.023)	.96 (.042)
2	-.04	29.5	30	.95 (.055)	.97 (.021)	.98 (.018)	.98 (.011)	.98 (.026)	.98 (.016)
		47.5	50	.97 (.032)	.98 (.010)	.96 (.022)	.97 (.028)	.95 (.020)	.97 (.025)
		10	10	.81 (.243)	.76 (.250)	.94 (.092)	.98 (.009)	.98 (.031)	.96 (.043)
3	-.05	27.7	30	.89 (.095)	.90 (.119)	.93 (.051)	.92 (.057)	.91 (.047)	.93 (.037)
		41.1	50	.78 (.081)	.84 (.101)	.81 (.056)	.79 (.085)	.82 (.038)	.82 (.039)
		9	10	.73 (.310)	.69 (.341)	.85 (.093)	.84 (.232)	.91 (.058)	.93 (.085)
4	-.03	23.1	30	.80 (.107)	.75 (.103)	.77 (.071)	.82 (.114)	.77 (.060)	.77 (.077)
		36.8	50	.74 (.048)	.78 (.071)	.74 (.041)	.75 (.051)	.74 (.029)	.76 (.049)
		4.7	10	.55 (.282)	.50 (.292)	.46 (.159)	.58 (.216)	.46 (.073)	.43 (.118)
5	-.01	14.3	30	.49 (.128)	.47 (.156)	.49 (.046)	.46 (.088)	.48 (.043)	.48 (.046)
		24.7	50	.46 (.066)	.47 (.072)	.48 (.043)	.48 (.056)	.50 (.037)	.49 (.048)
		8.6	10	.83 (.232)	.81 (.172)	.88 (.077)	.92 (.078)	.88 (.076)	.84 (.092)
6	-.04	22.3	30	.76 (.062)	.73 (.163)	.72 (.086)	.77 (.085)	.75 (.033)	.74 (.033)
		35.8	50	.69 (.037)	.76 (.081)	.72 (.060)	.72 (.051)	.72 (.032)	.74 (.041)
		2.8	10	.31 (.184)	.45 (.309)	.34 (.199)	.70 (.259)	.27 (.125)	.62 (.289)
7	-.05	3.7	30	.12 (.101)	.42 (.310)	.13 (.089)	.48 (.314)	.13 (.052)	.56 (.249)
		3.2	50	.06 (.066)	.62 (.238)	.06 (.025)	.62 (.265)	.06 (.035)	.63 (.273)

Table 2: Average posterior modes $\hat{r}$ and $\hat{q}$ for all scenarios, $\epsilon \in \{.5, 1, 2\}$ and $M \in \{10, 30, 50\}$. $\hat{\beta}_j$ provides the coefficient estimates from overall regression models. S_1 reports the average number of coefficient estimates below the threshold. Standard errors for the posterior modes $\hat{r}$ and $\hat{q}$ are in parentheses. Averages and standard errors are computed over ten queries.

so does $\hat{q}_{\texttt{NA}}$. In fact, for $M = 50$, we observe $\hat{q}_{\texttt{NA}}$ around .86, which are fairly large values. Thus, $\hat{q}_{\texttt{NA}}$ fulfills its role of adequately informing about the estimability of the coefficient. Regarding standard errors, we observe some large values particularly when $\epsilon = .5$ or $M = 10$. However, most of the times, the size of standard errors decreases as the value of ϵ and M increases.

Results reported in Tables 2 and 3 show how the performance of the considered verification measures can vary as a function of ϵ, M, n, and n_j. When the values of ϵ, M, n, and n_j are not small, the binomial and multinomial-based strategies provide accurate verifications. As a general guideline for ϵ and M, we recommend not to use small values such as 0.5 and 10, respectively. However, these analyses also show that even when $M > 10$, aspects of the dataset other than n and n_j can influence the performance of the verification measures. Having a synthetic dataset of the Status File can help find reasonable values for ϵ and M. With a synthetic dataset, analysts can simply compute, for different values of ϵ and M, the verification measures as many times as needed. Thus, if an analyst can find values of ϵ and M that work well with the synthetic dataset, then she can use those same values to submit a query and ask for a given verification measure based on the Status File. Finally, the multinomial based strategy gives analysts an extra tool to judge whether or not to rely on the obtained verification. Specifically, analysts

E	n	n/n_j	S_1	S_0	$S_{\texttt{NA}}$	M	$q_{\texttt{NA}}$ $\epsilon = .5$	$\epsilon = 1$	$\epsilon = 2$
			0	10	0	10	.06 (.074)	.08 (.092)	.04 (.045)
1	193144	.34	0	30	0	30	.04 (.064)	.03 (.034)	.01 (.011)
			.3	49.7	0	50	.01 (.018)	.02 (.028)	.01 (.005)
			10	0	0	10	.05 (.050)	.03 (.039)	.03 (.038)
2	305311	.14	29.5	.5	0	30	.03 (.022)	.01 (.016)	.02 (.014)
			47.5	2.5	0	50	.04 (.040)	.01 (.006)	.01 (.012)
			10	0	0	10	.03 (.015)	.04 (.051)	.05 (.048)
3	458114	.05	27.7	2.3	0	30	.01 (.009)	.02 (.015)	.01 (.012)
			41.1	8.9	0	50	.02 (.020)	.01 (.012)	.01 (.011)
			9	1	0	10	.18 (.218)	.04 (.024)	.05 (.082)
4	165210	.02	23.1	6.9	0	30	.04 (.053)	.05 (.047)	.01 (.004)
			36.8	13.2	0	50	.02 (.033)	.01 (.004)	.01 (.016)
			4.7	5.3	0	10	.06 (.049)	.06 (.082)	.01 (.005)
5	34154	.11	14.3	15.7	0	30	.03 (.060)	.02 (.016)	.01 (.021)
			24.7	25.3	0	50	.07 (.071)	.01 (.010)	.01 (.013)
			8.6	1.4	0	10	.12 (.134)	.06 (.091)	.04 (.036)
6	42077	.03	22.3	7.7	0	30	.09 (.085)	.04 (.053)	.02 (.021)
			35.8	13.7	.5	50	.07 (.102)	.03 (.029)	.01 (.014)
			2.8	2.3	4.9	10	.24 (.232)	.44 (.171)	.41 (.132)
7	705	.05	3.7	3.1	23.2	30	.75 (.089)	.75 (.047)	.75 (.066)
			3.2	2.2	44.6	50	.85 (.061)	.87 (.041)	.87 (.031)

Table 3: Average posterior modes $\hat{q}_{\texttt{NA}}$ for all scenarios, $\epsilon \in \{.5, 1, 2\}$ and $M \in \{10, 30, 50\}$. S_1 and S_0 report the average number of coefficient estimates below and above the threshold, respectively. $S_{\texttt{NA}}$ displays the average number of nonestimable coefficients. Standard errors for the posterior modes $\hat{q}_{\texttt{NA}}$ are in parentheses. Averages and standard errors are computed over ten queries.

should refrain from using $\hat{q}$ when $\hat{q}_{\texttt{NA}}$ is large. We hope these general guidelines aid analysts to more easily use and interpret the verification measures.

References

K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84. ACM, 2007.