An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices

John M. Abowd and Ian M. Schmutte

Online Appendix

August 14, 2018

A Potential Database Reconstruction Attack against the 2010 Decennial Census

In Section I.B.3 we discuss the potential for a database reconstruction attack against the decennial census based on the large number of summary tables published from the confidential micro-data. Using the schema in the public documentation for PL94-171, Summary File 1, Summary File 2, and the Public-use Micro-data Sample, and summarizing from the published tables, there were at least 2.8 billion *linearly independent* statistics in PL94-171, 2.8 billion in the balance of SF1, 2.1 billion in SF2, and 31 million in the PUMS https://www.census.gov/prod/www/decennial.html (cited on March 17, 2018). For the 2010 Census, the national sample space at the person level has approximately 500,000 cells. The unrestricted sample space at the census block level has approximately $500{,}000 \times 10^7$ cells. It might seem there are orders of magnitude more unknowns than equations in the system used for reconstruction. However, traditional statistical disclosure limitation (SDL) does not protect sample zeros. Consequently, every zero in a block, tract, or county-level table rules out all record images in the sample space that could have populated that cell, dramatically reducing the number of unknowns in the relevant equation system.

The deliberate preservation of sample zeros can be inferred from the technical documentation: "Data swapping is a method of disclosure avoidance designed to protect confidentiality in tables of frequency data (the number or percentage of the population with certain characteristics). Data swapping is done by editing the source data or exchanging records for a sample of cases. A sample of households is selected and matched on a set of selected key variables with households in neighboring geographic areas (geographic areas with a small population) that

have similar characteristics (same number of adults, same number of children, etc.). Because the swap often occurs within a geographic area with a small population, there is no effect on the marginal totals for the geographic area with a small population or for totals that include data from multiple geographic areas with small populations. Because of data swapping, users should not assume that tables with cells having a value of one or two reveal information about specific individuals" (U.S. Census Bureau 2012, p. 7-6).

B Randomized Response Details

A custodian collects data from a population of individuals, $i \in \{1, \ldots, N\}$. Each member of the population has a sensitive characteristic and an innocuous characteristic. The sensitive characteristic is $x_i = Y_i(1) \in \{0, 1\}$, with population proportion $\Pr[Y_i(1) = 1] = \pi$. This proportion, π, is the unknown population quantity of interest. The non-sensitive characteristic is $z_i = Y_i(0) \in \{0, 1\}$ with known population proportion $\Pr[Y_i(0) = 1] = \mu$. The custodian collects and publishes a mixture

$$d_i = T_i Y_i(1) + (1 - T_i) Y_i(0), \tag{B-1}$$

where T_i indicates whether the sensitive or the non-sensitive question was collected, with $\Pr[T_i = 1] = \rho$. The responses are independent of which information is collected: $(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i$. We also require that the non-sensitive item be independent of the sensitive item. This is not restrictive, since the innocuous question can literally be "flip a coin and report whether it came up heads," as in the original application.

The indicator T_i is not observed. Any data analyst observes only the reported variable d_i. However, as in a randomized controlled trial, the probability of T_i, ρ, is known with certainty. Furthermore, the analyst also knows the probability of the non-sensitive response, μ.

Define $\widehat{\beta} = \frac{1}{N} \sum_i d_i$, the empirical mean proportion of responses of one. Independence of T_i implies $\mathrm{E}\left[\widehat{\beta}\right] = \pi\rho + \mu(1-\rho)$. It follows that $\widehat{\pi} = \frac{\widehat{\beta} - \mu(1-\rho)}{\rho}$ is an unbiased estimator of π with variance $\mathrm{Var}[\widehat{\pi}] = \mathrm{Var}[\widehat{\beta}]\rho^{-2}$.

B1 Privacy under Randomized Response

For given ε, differential privacy requires both $\Pr[d_i = 1 | Y_i(1) = 1] \leq e^{\varepsilon} \Pr[d_i = 1 | Y_i(1) = 0]$, and $\Pr[d_i = 0 | Y_i(1) = 0] \leq e^{\varepsilon} \Pr[d_i = 0 | Y_i(1) = 1]$. Together, these expressions bound

the Bayes factor, which limits how much can learned about the sensitive characteristic upon observation of the collected response.

Making substitutions based on the data-generating model,

$$1 + \frac{\rho}{(1-\rho)\mu} \leq e^{\varepsilon} \tag{B-2}$$

and

$$1 + \frac{\rho}{(1-\rho)(1-\mu)} \leq e^{\varepsilon}. \tag{B-3}$$

For a given choice of μ, the differential privacy guaranteed by randomized response is the maximum of the values of the left-hand sides of equations (B-2) and (B-3). Hence, privacy loss is minimized when $\mu = \frac{1}{2}$. This is the case we will consider throughout the remaining discussion. We note doing so assumes that inferences about affirmative and negative responses are equally sensitive, which may not always be the case. The results of our analysis do not depend on this assumption. [57]

For randomized response, the differential privacy guarantee as a function of ρ is:

$$\varepsilon(\rho) = \log\left(1 + \frac{2\rho}{(1-\rho)}\right) = \log\left(\frac{1+\rho}{1-\rho}\right), \tag{B-4}$$

which follows from setting $\mu = \frac{1}{2}$ in equations (B-2) and (B-3).

[57] These observations highlight another allocation problem: how to trade off protection of affirmative responses for the sensitive item $Y_i(1) = 1$ against protection of negative responses $Y_i(1) = 0$. What do we mean? If ρ is fixed, then increasing μ reduces the Bayes factor in (B-2) (increasing privacy) and increases the Bayes factor in (B-3) (decreasing privacy). The underlying intuition is fairly simple. Suppose the sensitive question is "did you lie on your taxes last year?" Most tax evaders would prefer that their answer not be made public, but non-evaders are probably happy to let the world know they did not cheat on their taxes. In such a setting, with ρ fixed, we can maximize privacy for the tax evader by setting μ to 1. Recall μ is the probability of a positive response on the non-sensitive item ($Y_i(0) = 1$). If $\mu = 1$, then when the data report $d_i = 0, 1$ we know with certainty that $Y_i(1) = 0$ (i.e., i did not cheat on her taxes). In this special case, the mechanism provides no privacy against inference regarding non-evasion, but maximum attainable privacy (given the mechanism) against inference regarding evasion. This is the role the Bloom filter plays in the full RAPPOR implementation of randomized response (Erlingsson, Pihur and Korolova 2014). More generally, the choice of μ can be tuned to provide relatively more or less privacy against one inference or the other.

B2 Statistical Accuracy under Randomized Response

Expressed as a function of ρ, we denote the estimated share of the population with the sensitive characteristic

$$\widehat{\pi}(\rho) = \frac{\widehat{\beta}(\rho) - \mu(1-\rho)}{\rho} \tag{B-5}$$

where $\widehat{\beta}(\rho)$ is the population average response when the sensitive question is asked with probability ρ. Clearly,

$$\mathrm{E}[\widehat{\beta}(\rho)] = [\rho\pi + (1-\rho)\mu] \tag{B-6}$$

and

$$\mathrm{Var}[\widehat{\beta}(\rho)] = \frac{[\rho(\pi-\mu)+\mu](1-\rho(\pi-\mu)-\mu)}{N}. \tag{B-7}$$

It follows that

$$\mathrm{Var}[\widehat{\pi}(\rho)] = \frac{\mathrm{Var}[\widehat{\beta}(\rho)]}{\rho^2} = \frac{[\rho(\pi-\mu)+\mu](1-\rho(\pi-\mu)-\mu)}{\rho^2 N}. \tag{B-8}$$

We can define data quality as:

$$I(\rho) = \mathrm{Var}[\widehat{\pi}(1)] - \mathrm{Var}[\widehat{\pi}(\rho)]. \tag{B-9}$$

This measures the deviation in the sampling variance for the predicted population parameter, π, relative to the case where there is no privacy protection ($\rho = 1$).

B3 The Accuracy Cost of Enhanced Privacy under Randomized Response

Equations (B-4) and (B-9) implicitly define a functional relationship between data privacy, parameterized by ε, and accuracy, parameterized as I. This function tells us the marginal cost borne by individuals in the database necessary to achieve an increase in accuracy of the published statistics. We can characterize the relationship between accuracy, I, and privacy loss, ε, analytically. First, we invert equation (B-4) to get ρ as a function of ε:

$$\rho(\varepsilon) = \frac{e^{\varepsilon} - 1}{1 + e^{\varepsilon}}. \tag{B-10}$$

Next, we differentiate I with respect to ε via the chain rule: $\frac{dI}{d\varepsilon} = I'(\rho(\varepsilon))\rho'(\varepsilon)$:

$$I'(\rho) = \frac{2\,\mathrm{Var}[\widehat{\beta}(\rho)]}{\rho} - \frac{(\pi - \frac{1}{2})(1 - 2\pi)}{N^2\rho}. \tag{B-11}$$

and

$$\rho'(\varepsilon) = \frac{2e^{\varepsilon}}{(1 + e^{\varepsilon})^2} = \frac{1}{1 + cosh(\varepsilon)}. \tag{B-12}$$

Both derivatives are positive, so it follows that $\frac{dI}{d\varepsilon} > 0$. A similar derivation shows that $\frac{d^2I}{d\varepsilon^2} < 0$. Increasing published accuracy requires an increase in privacy loss at a rate given by $\frac{dI}{d\varepsilon} > 0$. Furthermore, achieving a given increment in accuracy requires increasingly large privacy losses.

C Details of the Matrix Mechanism

For a single query, we defined the ℓ_1 sensitivity in Definition 1. The results in Theorems 1 and 2 are defined in terms of the sensitivity of a workload of linear queries, which we denote ΔQ. Following Li et al. (2015),

Theorem 3 (ℓ_1 Query Matrix Sensitivity) Define the ℓ_1 sensitivity of Q by

$$\Delta Q = \max_{x,y \in \mathbb{Z}^{*|x|}, \|x-y\|_1 \leq 1} \|Qx - Qy\|_1 .$$

This is equivalent to

$$\Delta Q = \max_k \|q_k\|_1 ,$$

where q_k are the columns of Q.

For the proof, see Li et al. (2015, prop. 4).

D Details of Privacy Semantics

We provide technical definitions associated with the derivations in Kifer and Machanavajjhala (2012) described in Section IV.A.

Assume a latent population of individuals $h_i \in \mathcal{H}$ of size N^*. The confidential database, D, is a random selection of $N < N^*$ individuals, drawn independently from $\mathcal{H}$. In this context N is a random variable, too. The database also records

characteristics of each individual, which are drawn from the data domain χ. Denote the record of individual i as r_i. The event "the record r_i is included in database D" has probability π_i. Denote the conditional probability of the event "the record $r_i = \chi_a \in \chi$" given that r_i is in D as $f_i(r_i)$. Then, the data generating process is parameterized by $\theta = \{\pi_1, ..., \pi_N, f_1, ..., f_N\}$. The probability of database D, given θ, is

$$Pr[D \mid \theta] = \prod_{h_i \in D} f_i(r_i)\pi_i \prod_{h_j \notin D} (1 - \pi_j). \tag{D-13}$$

The complete set of paired hypotheses that differential privacy protects is

$$\mathcal{S}_{pairs} = \{(s_i, s_i') : h_i \in \mathcal{H}, \chi_a \in \chi\}, \tag{D-14}$$

where s and s' are defined in Section IV.A. By construction $\mathcal{S}_{pairs}$ contains every pair of hypotheses that constitute a potential disclosure; that is, whether any individual h_i from the latent population is in or out of the database D and, if in D, has record r_i.

E Derivation of the Data Utility Model

Recall that the matrix mechanism publishes a vector of answers, $M(x, Q)$ to the known set of queries, Q given an underlying data histogram x. The matrix mechanism is implemented by using a data independent mechanism to answer a set of queries represented by the query strategy matrix, A with sensitivity ΔA and pseudo-inverse A^+. Following Theorem 2, $M(x, Q) = Qx + Q(\Delta A)A^+e$ where e is a vector of *iid* random variables with $\mathbb{E}[e] = 0$ and whose distribution is independent of x, Q, and A. In what follows, we use the notation σ_e^2 to denote the common (scalar) variance of the elements of e. For example, when e is a vector of Laplace random variables with scale ε^{-1}, we know that $\sigma_e^2 = 2\varepsilon^{-2}$. Note that the variance of the vector e is $\mathbb{E}\left[ee^T\right] = \sigma_e^2\mathbb{I}$ where $\mathbb{I}$ is the identity matrix conformable with e.

Let $W_i = \Pi_i^T M(x, Q)$ be a person-specific linear function by which published statistics are transformed into wealth (or consumption). Individuals have utility of wealth given by a twice-differentiable and strictly concave function, $U_i(W_i)$. The total realized ex post wealth for i is $W_i = \Pi_i^T Qx + \Pi_i^T QA^+(\Delta A)e$. We assume i knows Q and the details of the mechanism M. Uncertainty is over x and e.

For notational convenience, we define a function $w_i(e; x) = \Pi_i^T Qx + \Pi_i^T QA^+(\Delta A)e$. Conditional on x, the expected utility of i from receiving the mechanism output is $\mathbb{E}_{e|x}[U_i(w_i(e; x))|x]$. We approximate this by taking expectations of a second-

order Taylor Series expansion of $h_i(e;x) = U_i(w_i(e;x))$ with respect to e evaluated at $e_0 = 0$.

Let $\nabla h_i(e_0;x)$ denote the gradient of h with respect to e and let $H_i(e_0;x)$ denote the Hessian. The second-order Taylor series expansion of $h_i(e;x)$ evaluated at e_0 is

$$h_i(e;x) \approx h_i(e_0;x) + (e-e_0)^T \nabla h_i(e_0;x) + \frac{1}{2!}(e-e_0)^T H_i(e_0;x)(e-e_0). \tag{E-15}$$

The gradient of h is

$$\nabla h_i(e_0;x) = U_i'(w_i(e_0;x))\Delta A \left(\Pi_i^T Q A^+\right)^T. \tag{E-16}$$

The Hessian is

$$H_i(e_0;x) = U_i''(w_i(e_0;x))(\Delta A)^2 \left(\Pi_i^T Q A^+\right)^T \left(\Pi_i^T Q A^+\right). \tag{E-17}$$

Note that we have used the chain rule in both derivations. We now evaluate the right hand side of equation (E-15) at $e_0 = 0$. Defining new notation, let $w_{i0}^x = w_i(0;x) = \Pi_i^T Q x$ and making substitutions for the gradient and Hessian, we have

$$h_i(e;x) \approx U_i(w_{i0}^x) + U_i'(w_{i0}^x)\Delta A \left[e^T \left(\Pi_i^T Q A^+\right)^T\right] + \frac{1}{2}U_i''(w_{i0}^x)\Delta A^2 \left[e^T \left(\Pi_i^T Q A^+\right)^T \left(\Pi_i^T Q A^+\right) e\right]. \tag{E-18}$$

Now, taking expections with respect to e, conditional on x

$$\mathbb{E}_{e|x}[h(e;x)|x] \approx U_i(w_{i0}^x) + \frac{1}{2}U_i''(w_{i0}^x)\Delta A^2 \cdot \mathbb{E}_{e|x}\left\{\left[e^T \left(\Pi_i^T Q A^+\right)^T \left(\Pi_i^T Q A^+\right) e\right]|x\right\}. \tag{E-19}$$

The first-order term drops out because $\mathbb{E}_{e|x}[e|x] = 0$ by assumption. Focusing on the quadratic form in the final summand, standard results imply

$$\begin{aligned}
\mathbb{E}_{e|x}\left\{\left[e^T \left(\Pi_i^T Q A^+\right)^T \left(\Pi_i^T Q A^+\right) e\right]|x\right\} &= tr\left[\mathbb{E}_{e|x}\left[ee^T|x\right]\left(\Pi_i^T Q A^+\right)^T \left(\Pi_i^T Q A^+\right)\right] && \text{(E-20)}\\
&= tr\left[\sigma_e^2 \mathbb{I}\left(\Pi_i^T Q A^+\right)^T \left(\Pi_i^T Q A^+\right)\right] && \text{(E-21)}\\
&= \sigma_e^2 tr\left[\left(\Pi_i^T Q A^+\right)^T \left(\Pi_i^T Q A^+\right)\right] && \text{(E-22)}\\
&= \sigma_e^2 ||\Pi_i^T Q A||_F^2. && \text{(E-23)}
\end{aligned}$$

The last expression is a basic property of the matrix Frobenius norm (Li et al. 2015).

Putting it all together, we have the following approximation to the expected utility for person i:

$$\mathbb{E}[U_i(W_i)] = \mathbb{E}_x\left[\mathbb{E}_{e|x}\left[h(e;x)|x\right]\right] \tag{E-24}$$

$$\approx \mathbb{E}_x\left[U_i\left(w_{i0}^x\right) + \frac{1}{2}U_i''(w_{i0}^x)\Delta A^2\sigma_e^2\|\Pi_i^T QA\|_F^2\right] \tag{E-25}$$

$$= \mathbb{E}_x\left[U_i\left(w_{i0}^x\right)\right] + \frac{1}{2}\mathbb{E}_x\left[U_i''(w_{i0}^x)\right]\Delta A^2\sigma_e^2\|\Pi_i^T QA\|_F^2. \tag{E-26}$$

Note that we have used the fact that A, Q, and Π_i^T are all independent of x.

From Theorem 2 the accuracy of the matrix mechanism is

$$I = -\sigma_e^2\left(\Delta A\right)^2\left\|QA^+\right\|_F^2. \tag{E-27}$$

We can therefore substitute accuracy, I, into the expression for expected utility

$$\mathbb{E}[U_i(W_i)] \approx \mathbb{E}_x\left[U_i\left(w_{i0}^x\right)\right] - \left\{\frac{1}{2}\mathbb{E}_x\left[U_i''(w_{i0}^x)\right]\frac{\|\Pi_i^T QA\|_F^2}{\|QA\|_F^2}\right\} \times I. \tag{E-28}$$

The expression above rationalizes a model for individual-specific data utility that is linear in accuracy, I: $v_i^{Data}(I) = a_i + b_i I$.

F Details of Legislative Redistricting Example

This appendix describes the legal background for the legislative redistricting example in Section VI.A. These properties of the SDL applied in the 2010 PL94-171 can be deduced from U.S. Census Bureau (2012, p. 7-6), as quoted in Appendix A, and the details provided in U.S. Census Bureau (2002), which also reveals that no privacy protection was given to the race and ethnicity tables in the 1990 redistricting data. The origin of the decision not to protect population and voting-age population counts is difficult to trace in the law. Public Law 105119, title II, 209, Nov. 26, 1997, 111 Stat. 2480, amended 13 U.S.C. Section 141 to provide that: "(h) ... In both the 2000 decennial census, and any dress rehearsal or other simulation made in preparation for the 2000 decennial census, the number of persons enumerated without using statistical methods must be publicly available for all levels of census geography which are being released by the Bureau of the Census for: (1) all data releases before January 1, 2001; (2) the data contained in the 2000 decennial census Public Law 94171 [amending this section] data file released

for use in redistricting; (3) the Summary Tabulation File One (STF1) for the 2000 decennial census; and (4) the official populations of the States transmitted from the Secretary of Commerce through the President to the Clerk of the House used to reapportion the districts of the House among the States as a result of the 2000 decennial census. ... . (k) This section shall apply in fiscal year 1998 and succeeding fiscal years." http://www.law.cornell.edu/uscode/text/13 13 U.S. Code (1954). These amendments to Title 13 concerned the use of sampling to adjust the population counts within states, as is permitted even under current law. They gave standing to obtain a copy of population count data that were not adjusted by sampling, should the Census Bureau publish such data, which it did not do in 2000 nor 2010. Even so, only the reapportionment of the House of Representatives must be done without sampling adjustments (U.S. Supreme Court 1999). Sampling aside, other statistical methods, like edits and imputations, including whole-person substitutions, are routinely applied to the confidential enumeration data before any tabulations are made, including those used to reapportion the House of Representatives. These methods were upheld in Utah v. Evans (U.S. Supreme Court 2002).

References

13 U.S. Code. 1954. "USC: Title 13 - Census Act."

Erlingsson, Úlfar, Vasyl Pihur, and Aleksandra Korolova. 2014. "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response." *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14*, 1054–1067.

Kifer, Daniel, and Ashwin Machanavajjhala. 2012. "A rigorous and customizable framework for privacy." *Proceedings of the 31st symposium on Principles of Database Systems - PODS '12*, 77.

Li, Chao, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. 2015. "The matrix mechanism: optimizing linear counting queries under differential privacy." *The VLDB Journal*, 24(6): 757–781.

U.S. Census Bureau. 2002. "Census Confidentiality and Privacy 1790 to 2002." *https://www.census.gov/prod/2003pubs/conmono2.pdf*, (Cited on March 22, 2018).

U.S. Census Bureau. 2012. "2010 Census Summary File1–Technical Documentation." Department of Commerce, Economics and Statistics Administration.

U.S. Supreme Court. 1999. "DEPARTMENT OF COMMERCE v. UNITED STATES HOUSE (98-404) No. 98—404, 11 F. Supp. 2d 76, appeal dismissed; No. 98—564, 19 F. Supp. 2d 543, affirmed." https://www.law.cornell.edu/supct/html/98-404.ZO.html, (Cited on March 26, 2018).

U.S. Supreme Court. 2002. "UTAH V. EVANS (01-714) 536 U.S. 452 182 F. Supp. 2d 1165, affirmed." https://www.law.cornell.edu/supct/html/01-714.ZS.html, (Cited on March 22, 2018).