Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Bayesian differential privacy for small counties and individual commodities

by Balgobin Nandram, Habtamu Benecha and Linda J. Young

Release date: June 29, 2026

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Survey Methodology, June 2026
Vol. 52, No. 1, pp. 61-94
Statistics Canada, Catalogue No. 12-001-X

Bayesian differential privacy for small counties and individual commodities

Balgobin Nandram, Habtamu Benecha and Linda J. Young[1]

Abstract

We construct a hybrid Bayesian method, which includes a differentially private mechanism, to mask Census county totals for a U.S. state on acreage of a commodity. We use surrogates for data collected at the farm level from a past U.S. Census of Agriculture to illustrate our procedure. We use two Bayesian small area models (parametric and mixture) to accommodate the smaller counties with fewer farms and some counties with large acres. In these models, the Laplace distribution provides a differentially private mechanism. In pre-processing, we also incorporate the Census weights to form the observed total acreage, a scaling factor to the Laplace mechanism for each county, a square-root transformation of the observed total acreage to avoid negative masked estimates especially for small counties, and the p-percent rule and the 3+ rule to partition the counties into suppressed counties, non-sensitive counties and sensitive counties. Because of difficulties in specifying and tuning the privacy budget (an unknown parameter), to balance security and utility, we specify a prior for the privacy budget, where the values are not specified, and the Gibbs sampler is used to fit the hierarchical Bayesian models. In post-processing, we use Bayesian predictive inference to obtain masked county acreages, and this includes a benchmarking so that the masked state total matches the observed state total. As a measure of reliability of the Bayesian procedure, we use the posterior coefficients of variation for the masked posterior means of the counties. As a measure of utility, we use the absolute relative errors for the individual counties, together with other global measures. For the sensitive counties, there are some differences between the two small area models but both are much better than an individual area model; the mixture model being the best compromise for security and utility.

Key Words: Bayesian predictive inference; Benchmarking; Laplace distribution; Mixture model with stick-breaking weights; p-percent rule; Utility; Post-processing.

1. Introduction

It is likely that differential privacy (DP) will become important for government agencies. DP was used in the 2020 Census by the U.S. Census Bureau, but there are several challenges with its impact on data quality and usability. However, it is known that ϵ- differential privacy is the state-of-the-art *formal* privacy model for releasing sensitive information while protecting privacy (e.g., Williams and Bowen, 2023). Drechsler (2023) pointed out that despite the excitement in academia and industry, most agencies, with the prominent exception of the U.S. Census Bureau, have been reluctant to even consider the concept for their data release strategy. See also Reiter (2019) for discussion on DP and federal data releases. This suggests much more work needs to be done to make DP more practical and accessible. Yet, this is a difficult mathematical problem and conversion to statistics is a current research activity; see Dwork (2006), Dwork, McSherry, Nissim and Smith (2006) and Dwork, Naor, Reingold, Rothblum and Vadhan (2009) for the original formulation and the development in cryptography. We provide a hybrid Bayesian method, which includes a DP mechanism, to mask Census county totals for a U.S. state on acreage of a commodity.

1. Balgobin Nandram, Habtamu Benecha and Linda J. Young, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609. E-mails: balnan@wpi.edu, habtamu.benecha@usda.gov and linda@youngstatconsulting.com.

DP mechanisms work by injecting carefully calibrated noise (or randomness) into the observed confidential data to protect the privacy of individual respondents (e.g., farmers). Utility refers to the usefulness of the noisy data (or statistical results) after any privacy preserving mechanism is applied. Utility exists in a fundamental trade-off with privacy, and these two must be carefully balanced. We want the relative difference between the observed confidential data and the masked data to be small, but at the same time we strictly cannot reveal the identity of the farmers.

Many government agencies, including the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA), use cell suppression (both primary and secondary) to protect confidentiality of respondents (see Cox, 1980, 1995 and Cox, Kelly and Patil, 2004). This approach can be used to protect both frequency data and continuous data. In contrast, our proposed disclosure control method applies random noise to sensitive cells, obtained from possibly a differentially private mechanism, so that published tables can have all the cells populated. We analyze simulated harvested acres (an economic variable); other variables can be studied the same way or may be linked to harvested acres.

Duncan and Lambert (1986, 1989), although not fully related to our work, developed a decision-theoretic framework for risk assessment that includes an intruder's objectives and strategy for compromising the database and the information gained by the intruder is developed. They described two kinds of microdata disclosure, distinguished-disclosure of a respondent's identity and disclosure of a respondent's attributes as a result of an unauthorized identification. They also presented a formula for the risk of identity disclosure, and evaluated a simple approximation.

Rubin (1993) proposed to release only synthetic microdata, constructed using multiple imputation. The objective is to produce microdata that can be analyzed using standard statistical software. This method is followed closely by Raghunathan, Reiter and Rubin (2003). The basic procedure is to simulate multiple copies of the population from which these respondents have been selected and release a random sample from each of these synthetic populations. Users can analyze the synthetic sample data sets with standard complete-data software for simple random samples, then obtain valid inferences by combining the point and variance estimates using the methods they provide; see Little (1993). Multiple copies of the original data set can also be generated using DP. We cannot release microdata (e.g., farm-level data), but we can release county-level data after a statistical disclosure control is performed.

Bowen and Liu (2020) examined current differentially private data synthesis techniques for releasing individual-level surrogate data instead of the original data, compared the techniques conceptually and evaluated the statistical utility and inferential properties of the synthetic data via each technique through extensive simulation studies. However, there are no discussions within the Bayesian framework; their review consists solely of current non-Bayesian techniques. However, Hu and Bowen (2024) described Bayesian methods in Section 3 of their paper and they compared Bayesian and non-Bayesian models as well as parametric and non-parametric models (including the Dirichlet process); they did not cover Bayesian methods in differential privacy. Essentially, they used regular Bayesian models (e.g., regression models) that reflect the main features of the confidential data, and posterior predictive inference was used for

masking. The Laplace distribution, a randomization mechanism, is an important component in our Bayesian models (i.e., differential privacy is used to form a hybrid method).

Bayesian methods for differential privacy are under rapid development. For example, Zhang, Rubinstein and Dimitrakakis (2016) studied how to communicate findings of Bayesian inference to third parties, while preserving the strong guarantee of differential privacy. Their main contributions are on four different algorithms for private Bayesian inference with probabilistic graphical models. The Laplace distribution is one of the mechanisms they have used to study differential privacy. They have also looked at binary data, but they have used a non-negative Laplace distribution, truncated on the upper end at the sample size. Differentially private inference for binomial data is useful in its own right; see Awan and Slavkovic (2019). Hu, Williams and Savitsky (2025) developed a pseudo Bayesian synthesizer and discussed the utility-risk trade-off. A fully Bayesian synthesizer is difficult to obtain within their framework. At the U.S. Census Bureau (frequency data), Janicki, Holan, Irimata, Livsey and Raim (2024) described a Bayesian method to produce a set of noisy measurements for tabular frequency data, and post-processed (main contribution) these noisy measurements using a model through posterior inference. This method incorporates known constraints for counts and ratios, and they showed improved inference of the partially de-noised data. They have confidential data, Y, and they want to publicly provide masked data. They constructed $Z = Y + E$, where E is the noise (a standard non-Bayesian analysis), and they used the Bayesian model, $Z \mid Y, \boldsymbol{\theta}$ to add the constraints to Y and to update Y.

Rinott, O'Keefe, Shlomo and Skinner (2018) reviewed differential privacy in the dissemination of frequency tables. They studied confidentiality protection for perturbed frequency tables, including the trade-off with analytical utility. However, we are not concerned with frequency tables in this paper, and our contribution is to develop a differential private methodology on masking a commodity with the intention to provide a more extensive methodology on linked and hierarchical tables for the U.S. Census of Agriculture.

We note that there are other ways harvested acres can be masked, already obtained from the Census of Agriculture, for release to the public. For example, a government agency can use convolution instead of differential privacy. One can multiply the harvested acres by $U \sim$ Uniform$(1-a, 1+a)$, where a is near 0, say $a = 0.10$, choosing a to get reasonable security and utility; see Evans, Zayata and Slanta (1996). This idea is related to randomized response (a differential privacy mechanism) in sample surveys, when a variable, which must be secured, is convoluted (e.g., multiplied) with another variable using a random mechanism. See Poole (1974) for pioneering work on randomized response for continuous variables in survey sampling; also, see Nandram and Yu (2019) for a review on discrete data. Again, this is not the direction of our research. The standard approach is to use cell suppression, but it is desirable to explore the useful features of differential privacy, although difficult to implement in Bayesian models. Our methodology provides a general framework for many problems with more complex data structures (e.g., multi-way tables, linked tables and hierarchical tables); of course, the models will be different.

We note that the acres for each farm are weighted up using Census weights. These weights are estimated for each responding Census farm by using a dual system estimation (DSE) approach to adjust for under-coverage in the Census list frame as well as non-response and mis-classification of farms. After calibration,

the weights are summed for counties or states to obtain the number of farms at that geographical level. The weighted sum of reported values of a commodity over a geographical area of interest gives the total estimate of that commodity. [It is worth noting that the USDA defines a farm as any place that produced or sold, or normally would have sold, at least one thousand dollars of agricultural products during a year.]

Simulated data are used for all applications in this paper. The harvested acreages of a commodity are simulated for a hypothetical state with 75 counties; each county has a number of farms, and so these are microdata. We will refer to the commodity as "COM", the state as "STAT", and the original confidential table (of acreages by county and state) by "TAB". Here, TAB has a total of 76 cells including the state total; the state total is not confidential. Interest is in protecting cells of the table by adding appropriate noise to the cell values.

In this paper, we consider COM for all counties of STAT. Let n_i denote the number of farms in the i^{th} county. Let $(w_{ij}, y_{ij}^*), j = 1,\ldots,n_i, i = 1,\ldots,\ell,$ denote the Census weights and the observed harvested acres associated with the j^{th} farm within the i^{th} county. Then $y_{ij} = w_{ij} y_{ij}^*$ is the Census harvested acres associated with the w_{ij} farms represented by the observed j^{th} farm within the i^{th} county. Although we do not use covariates, we describe how we can include them in the concluding section. In a future paper, we will look at two-way tables (county by commodity) with several counties and commodities (not just one) of acres.

We use the Laplace mechanism to provide differential privacy to TAB, coupled with additional assumptions in the hierarchical Bayesian models to reflect the main features of the data. With the use of differential privacy and a small budget, $\epsilon,$ it may be safe to release data, and this will depend on the agency. The main issue is that utility might not be acceptable. Therefore, our method is a hybrid Bayesian procedure because it is not based solely on differential privacy. We use small area models because farms within a county, and counties within a state, can have similar practices. Also, some counties can have sparse data, and this can become problematic. As a secondary objective, we will show the usefulness of the small area models by comparing them with a single area (i.e., individual county) model.

We assume a population model (without noise) and a sample model (with Laplace noise) is derived from it. The population model is adjusted by adding Laplace noise to it, and prediction is done using the population model after the sample model is fit. We use a basic model to motivate our two hierarchical Bayesian models which are fully described in Section 3. The basic model is

$$y_{ij} \mid \mu_i, \sigma^2 \overset{\text{ind}}{\sim} \text{Normal}(\mu_i, \sigma^2), j = 1,\ldots,n_i, i = 1,\ldots,\ell.$$

This model is used for prediction of the y_{ij} to obtain the county totals after noise is added to the μ_i and σ^2 using a differentially private mechanism. While this model is not robust to non-normality, the mixing effect of the Laplace distribution will make it robust against non-normality to some degree in the sample model.

The sample model is formed by adjusting our population model to accommodate the observed confidential data as

$$y_{ij} \mid \mu_i, r_i, t_i, \sigma^2 \overset{\text{ind}}{\sim} \text{Normal}\left\{\mu_i + (2r_i - 1)\, t_i, \sigma^2\right\}, j = 1, \ldots, n_i, i = 1, \ldots, \ell,$$

where $(2r_i - 1)\, t_i$ is a convenient representation of the Laplace random variable, which is probabilistically constructed as

$$r_i \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{1}{2}\right), \quad t_i \mid \epsilon \overset{\text{ind}}{\sim} \text{Gamma}\,(1, \epsilon)$$

(same as the exponential density), and ϵ is the privacy budget with the $(2r_i - 1)\, t_i$ having mean zero; see Appendix A for the definition of differential privacy. Note that $f(t_i \mid \epsilon) = \epsilon e^{-\epsilon t_i}$, $t_i > 0$ with $E(t_i) = 1/\epsilon =$ SD (t_i). Thus, smaller ϵ means larger variability, and less risk (more security) and less utility. When ϵ is selected to be small (e.g., $\epsilon =$ 1, 2), the differential privacy procedure is meant to give small risk, but not necessarily large utility. The trade-off in risk and utility cannot be avoided.

There is further robustness against non-normality when priors are specified for the μ_i, σ^2 and ϵ; see Section 2 for the prior on ϵ and Section 3 for the two hierarchical Bayesian models. The hierarchical Bayesian models, particularly the mixture model, induce robustness through mixing.

Note that the model is fit to microdata (farm-level data) to ensure robust protection of individual-level information. DP is used to add noise to all sensitive counties simultaneously, not one at a time. That is, the data set consists of all farms in all counties. The definition of differential privacy applies to all relevant counties at once. Therefore, we are constructing a novel hybrid method that incorporates a differentially private mechanism into a Bayesian model.

It is worth noting that the joint probability density function of all the observed data, $\mathbf{y}$, is

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{\ell} n_i \left\{\tilde{s}_i^2 + \left(\bar{y}_i - \mu_i - (2r_i - 1)\, t_i\right)^2\right\}\right\},$$

where $\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, $\tilde{s}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2, i = 1, \ldots, \ell.$ It is good that $f(\mathbf{y} \mid \boldsymbol{\mu}, \sigma^2)$ is not a function of the individual farm-level data, and clearly this is true for any posterior density with this likelihood. We also note that this is not directly a function of the county totals (i.e., it is a function of $\tilde{s}_i^2$ and $(\bar{y}_i - \mu_i - (2r_i - 1)\, t_i)^2$). In the joint posterior density, which contains all information, the $\tilde{s}_i^2$ and $\bar{y}_i$ are held fixed.

We use a small area model with the square-root transformation on the farm acres. This model allows the masked data to be positive under back transformation, satisfying the positivity constraint. It is possible to use inequality constraints; see for example Nandram, Cruze and Erciulescu (2023), but it is simpler to use the square-root transformation, which also helps with robustness to non-normality. Our masking procedure has the following steps.

a. The sample model is fit to the square root of observed data to maintain the positivity constraint. Noise is added to this model via the Laplace differential privacy mechanism.

b. We use Bayesian predictive inference to obtain the masked data. Once the sample model is fit, although there are several choices, Bayesian predictive inference is done using the population model in post-processing.

c. In an output analysis, the masked data are benchmarked to the observed data by raking. That is, the masked data from the counties are benchmarked so that the masked state total is the same as the observed state total. This is done at the post-processing stage. Note that the observed state total is not part of the masking procedure (model fitting).

This procedure will produce masked data with relatively increased precision over the Bayesian predictive inference in (b) because benchmarking in (c) will increase precision. This benchmarking procedure will not affect privacy because the observed farm acres are not used in the analysis (Dwork and Roth, 2014, Proposition 2.1). Williams and Bowen (2023) stated that post-processing provides opportunities to improve utility because available public or expert knowledge can be used to reduce the amount of noise without accruing additional privacy loss (post-processing theorem). They also said that more information should be incorporated into a differential private (DP) analysis, but as far as we and they know, many researchers are not doing so.

It is worth recalling and it is important to note that all results and demonstrations in this paper are based on simulated data created to emulate the structure and characteristics of the Census of Agriculture. This simulated dataset was generated exclusively for methodological illustration and does not contain any confidential or actual Census microdata. The approach and findings presented here are intended to showcase the feasibility and performance of the proposed methods in a realistic but fully simulated environment.

This paper has five sections, including this introductory section, and three appendices with some pertinent technical details. In Section 2, we provide some background information. Specifically, we have included a scale factor to the Laplace distribution, a feature so that the overall budget in the modeling is ϵ as would normally be specified in the definition of differential privacy (Williams and Bowen, 2023), and an assessment of sensitive counties with some discussions on local and global sensitivity. In Section 3, we describe two models, a parametric model and a robust mixture model (i.e., mixture model with stick-breaking weights) to accommodate large acres (outliers) and non-normality. In Section 4, we have a comprehensive comparison of the two models; also both models are compared with an individual area model. We also partially present masked data for TAB in STAT. In Section 5, we present concluding remarks. Appendix A has a brief description of the notion of ϵ-DP (differential privacy), Appendix B describes the p-percent sensitivity rule, and Appendix C describes a model that can be used for percentages to show greater generality.

2. Background information

We discuss the total budget; both the number of farms per county and a scaling factor needed in the Laplace distribution affect the total budget. We also discuss the predictive approach to mask the weighted Census data and the square-root transformation to provide positive masked county data.

Using the Laplace differential privacy mechanism, we apply a simple weighting scheme to obtain the total budget (see Appendix A) because sequential composition is inappropriate and parallel composition is

not available in small area estimation. We assume that only a single query is used, and therefore sequential composition is inappropriate. We are assuming that there are similarities among the counties in small area estimation, so counties are not really non-overlapping. Therefore, parallel composition is also inappropriate. See Williams and Bowen (2023) for a good discussion of the two composition theorems that are usually used in differential privacy. Let n_i denote the number of farms in the i^{th} county, $i=1,\ldots,\ell$. In order to move forward, we apply the simple weighting scheme to get an overall budget, which permits larger counties to have larger allocations of the budget. Because we want the overall budget to be ϵ, $a_o<\epsilon<b_o$, for $i=1,\ldots,\ell$, we define $a_{oi}=\frac{n_i}{\sum_{i'=1}n_{i'}}$, and therefore, we assume that the overall budget is $\sum_{i=1}^{\ell}a_{oi}\epsilon=\epsilon$.

We now consider the difficulty of scaling, which arises in the Bayesian differential privacy method. Observe that we are just adding $(2r_i-1)\,t_i$ to the mean county total, and there is a scaling problem with t_i; e.g., see Janicki et al. (2024). The scale is unity for all counties and this is not flexible enough, but total acreages over counties are variable; both scale and ϵ affect security and utility. One way to overcome this problem is to introduce a scaling factor, t_{oi},

$$t_i\,|\,\epsilon \overset{\text{ind}}{\sim} \text{Gamma}\left\{1,\frac{\epsilon}{t_{oi}}\right\}, i=1,\ldots,\ell,$$

where $\text{E}(t_i)=\frac{t_{oi}}{\epsilon}=\text{SD}(t_i)$ and t_{oi} are specified. Clearly, t_{oi} are not identifiable, and therefore, ϵ is not identifiable. One would need an alternative source of data to specify these scale parameters and perhaps past data, if available, can be used to obey the Bayesian paradigm.

For convenience, we use the weighted Census data, $y_{ij}, j=1,\ldots,n_i, i=1,\ldots,\ell,$ to obtain t_{oi} and we assume that they are known. We use the median absolute deviation (MAD), robust against outliers and normality. Letting $\tilde{y}_i=\text{ med}\{y_{ij}, j=1,\ldots,n_i\},$ we use

$$t_{oi}=1.4826\text{ med }\{|\,y_{ij}-\tilde{y}_i\,|, j=1,\ldots,n_i\}, i=1,\ldots,\ell,$$

a consistent estimator of the standard deviation. For those t_{oi} that are zeros, it is straight forward to make an adjustment. Clearly, other estimators (e.g., range, sample standard deviation) are possible, but these are heavily influenced by outliers; the inter-quartile range is a competitor. Therefore, to accommodate the overall budget of ϵ, we must incorporate both the sample sizes and the scaling factors. Both the sample sizes and the scaling factors affect risk and utility, and therefore, they must be incorporated into the model simultaneously. Then, for one possibility we can take

$$a_{oi}=\frac{n_i/t_{oi}}{\sum_{i'=1}^{\ell}n_{i'}/t_{oi'}},\ i=1,\ldots,\ell,$$

where $\sum_{i=1}^{\ell}a_{oi}=1$. This means that we need to update our assumption to

$$t_i\,|\,\epsilon \overset{\text{ind}}{\sim} \text{Gamma}\,\{1,a_{oi}\epsilon\}, i=1,\ldots,\ell.$$

However, note that if the scaling factors, $t_{oi}, i=1,\ldots,\ell,$ are the same, they will cancel from a_{oi} to get $a_{oi}=\frac{n_i}{\sum_{i'=1}^{\ell}n_{i'}}, i=1,\ldots,\ell$ (i.e., the scaling factors do not matter).

The data we actually have are $y_{ij}, j = 1,\ldots,n_i, i = 1,\ldots,\ell,$ where n_i is the number of farms under COM in each county and $y_{ij}, i = 1,\ldots,\ell,$ denote the harvested acres (weighted) in these farms. These are microdata. We can transform y_{ij} (acres are positive) so that when we back transform, we are on a positive scale (e.g., $s = \sqrt{y}$ and then $y = s^2$ is positive). We do not use the logarithmic transformation because when the logarithm is re-transformed, the Bayesian predictive distribution is not well defined; e.g., see Manandhar and Nandram (2021, 2025) who used generalized gamma distributions instead of the logarithmic transformation. Of course, other transformations that lead to positivity on the original scale are possible.

The good news is that, for small $\epsilon,$ the DP model is reliable with acceptable risk. Unfortunately, this depends on the choice of $\epsilon.$ Generally, utility is low in any DP model, and post-processing is necessary; see Williams and Bowen (2023) for a discussion of the post-processing theorem. As described, we do not need to specify any value of $\epsilon,$ and just a simple prior distribution is needed. Indeed, this is a novelty of the Bayesian paradigm, and a sensitivity study to the choice of ϵ is not needed. A mixture model is flexible and can be set up to deal with outliers. We note that Lee and Clifton (2011) used the observed data to derive an upper bound for ϵ to express uncertainty about $\epsilon,$ but we need to specify the probability a farmer's information is disclosed. Following Lee and Clifton (2011), Hsu, Gaboardi, Haeberlen, Khanna, Narayan, Pierce and Roth (2014) obtained a range for ϵ (both lower and upper bounds) based on a simple threat model, where economic variables are used rather than the actual study variable, but these economic variables are inappropriate in our application.

We have computed the local sensitivity of the observed data (county totals), which is based only on the observed county totals. It is simply the range of the data which is 3,498. Note that the three quartiles of the data are 60.5, 241.4 and 671.4 with an inter-quartile range (IQR) of 610.9 (i.e., the local sensitivity is 5.73 = 3,498/610.9 standard units). If the minimum and maximum values of all data sets are the same, this would have been global sensitivity, but this is not true; while one can set the minimum to be zero, the upper bound is unknown, but the global sensitivity cannot be infinite simply because acres must be bounded. Therefore, the global sensitivity must be at least (or approximately) 3,498 (clearly not infinity), and there are many counties, which are individually sensitive. We note that county totals are much more variable than county averages, and if county averages are used the local/global sensitivity would be a lot lower. We can obtain global sensitivity using a model-based analysis, but this is not necessary.

Clearly, any county with one or two farms is sensitive, and it is possible that a county with three farms is sensitive, so some government agencies use the 3+ rule, but there are other sensitivity rules as well; see Bring and Wang (2014) for a review. We need to identify counties with more than three farms that are sensitive, which is done using the p-percent sensitivity rule. Here, we are doing data analysis based on the p-percent sensitivity rule with $p = 10$ to illustrate our methodology. We define the p-percent sensitivity rule in Appendix B.

Finally, at the advice of NASS, we partition the counties into three sets. The first set consists of all the counties that satisfy the 3+ rule, where all counties with at most three farms are suppressed. This suppressed set of counties is eliminated from any modeling. The second set consists of all non-sensitive counties; noise

is not added to these counties. The third set consists of the sensitive counties, and noise is added to them using DP. The second and third sets are the non-suppressed counties and the partition of these counties is determined by the p-percent sensitivity rule; the small area models are applied to these non-suppressed counties, but as stated above noise is not added to the observations from non-sensitive (non-suppressed) counties.

Of the 75 counties, there are 17 suppressed counties (first set) that meet the 3+ rule, 38 non-sensitive counties (second set) and 20 sensitive counties (third set) that meet the p-percent rule. It is important to note that only the sensitive counties are benchmarked to the total of the observed acres for all these counties. While this aggregated total is expected to be sensitive (non-disclosable), the p-percent rule shows that it is not. This is done by imagining that all the farms in these sensitive counties come from a giant county (i.e., one sub-population). Benchmarking all non-suppressed counties to the observed non-suppressed county total is problematic because the non-sensitive counties tend to be larger than the sensitive counties, thereby creating a bias.

3. Bayesian models

We construct two Bayesian models, and we have used predictive inference to provide the final masked data. In Section 3.1, we use the Bayesian version of the one-way random effects model with a Laplace differential privacy mechanism. This is the parametric model discussed in Nandram, Toto and Choi (2011) for statistical analyses, not differential privacy. In Section 3.2, in order to accommodate outliers, we use a mixture model with a random number of components and random stick-breaking weights (i.e., generalized Dirichlet distribution); see Ishwaran and James (2001) for a comprehensive review of the stick-breaking priors. This is the mixture model, which should be robust against outliers in the farm acres and non-normality. In Section 3.3, we discuss a simple additional feature in our models to help mitigate over-shrinkage and how to perform Bayesian predictive inference to provide the masked data. It is worth noting that our models are not simple because they contain weakly identified parameters.

3.1 Parametric model

In our sample model, we assume

$$y_{ij} \overset{\text{ind}}{\sim} \text{Normal}\left\{\mu_i + (2r_i - 1)\, t_i, \sigma^2\right\},\ j = 1,\ldots,n_i,$$

$$r_i \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{1}{2}\right),\ t_i \overset{\text{ind}}{\sim} \text{Gamma}\left(1, a_{oi}\epsilon\right),\ \mu_i \overset{\text{ind}}{\sim} \text{Normal}\left(\theta, \frac{\rho}{1-\rho}\sigma^2\right), i = 1,\ldots,\ell,$$

where we prefer to use Gamma $\left(1, a_{oi}\epsilon\right)$ rather than Exponential $\left(a_{oi}\epsilon\right)$ as is usually done (these are the same), and the scaling factors, a_{oi}, are specified; see Section 2. Note that the μ_i are county effects,

$(2r_i - 1)\, t_i$ are the Laplace noise added to the county effects, and that we do not use any farm-level effects because these will not be identifiable. A priori,

$$\epsilon \sim \text{Gamma}(1,1), a_o < \epsilon < b_o;\ \pi(\theta, \sigma^2, \rho) \propto \frac{1}{\sigma^2},$$

where (a_o, b_o) are also specified to capture values of ϵ $(\epsilon = 1, 2),$ the values normally used in the definition of differential privacy; smaller values are preferred to give higher security but poorer utility. We have taken $a_o = 0,\ b_o = 2.25;\ a_o$ cannot be smaller than 0 and b_o is an upper bound. These specifications will give high security, and therefore, poor utility; later a utility-risk trade-off mechanism will be constructed.

Let $n = \sum_{i=1}^{\ell} n_i$ denote the total number of farms. Then, using Bayes' theorem, the joint posterior density is

$$\pi(\boldsymbol{\mu}, \mathbf{r}, \mathbf{t}, \epsilon, \theta, \sigma^2, \rho \mid \mathbf{y}) \propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \left(\frac{1-\rho}{\rho\sigma^2}\right)^{\ell/2} e^{-\epsilon} I_{\epsilon}(a_o, b_o) \times$$
$$\prod_{i=1}^{\ell} \left[a_{oi} \epsilon \exp\{-a_{oi} \epsilon t_i\} \prod_{j=1}^{n_i} \exp\left\{ -\frac{1}{2\sigma^2} \left\{ y_{ij} - \mu_i - (2r_i - 1)\, t_i \right\}^2 \right\} \right] \prod_{i=1}^{\ell} \exp\left\{ -\frac{1-\rho}{2\rho\sigma^2} (\mu_i - \theta)^2 \right\}.$$

Next, we present the conditional posterior densities (CPDs) required to run the Gibbs sampler.

Letting $\lambda_i = \frac{\rho n_i}{\rho n_i + 1 - \rho}, i = 1, \ldots, \ell,$ it is easy to show that the CPDs of the μ_i are

$$\mu_i \overset{\text{ind}}{\sim} \text{Normal}\left\{ \lambda_i \left(\bar{y}_i - \frac{(2r_i - 1)\, t_i}{n_i} \right) + (1 - \lambda_i)\, \theta,\ (1 - \lambda_i) \frac{\rho}{1-\rho} \sigma^2 \right\},\ i = 1, \ldots, \ell.$$

Letting

$$q_i = \prod_{j=1}^{n_i} \left\{ \frac{\phi\left(\frac{y_{ij} - \mu_i + t_i}{\sigma}\right)}{\phi\left(\frac{y_{ij} - \mu_i + t_i}{\sigma}\right) + \phi\left(\frac{y_{ij} - \mu_i - t_i}{\sigma}\right)} \right\},$$

where $\phi(\cdot)$ is the standard normal density, the CPDs of the r_i are

$$r_i \overset{\text{ind}}{\sim} \text{Bernoulli}(q_i), i = 1, \ldots, \ell.$$

It is also true that the CPD of ϵ is

$$\epsilon \sim \text{Gamma}\left(\ell + 1, \sum_{i=1}^{\ell} a_{oi} t_i + 1 \right) I_{\epsilon}(a_o, b_o).$$

The t_i are transformed to $t_i = \frac{\gamma_i}{1 - \gamma_i}, i = 1, \ldots, \ell,$ and the grid method is used to sample their CPDs (truncated normal densities can be used). We are left to sample the CPDs of $\theta, \sigma^2, \rho,$ and their joint CPD is

$$\pi(\theta, \sigma^2, \rho) \propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \left(\frac{1-\rho}{\rho\sigma^2}\right)^{\ell/2} \times \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \left\{ y_{ij} - \mu_i - (2r_i - 1) t_i \right\}^2 \right\} \exp\left\{ -\frac{1-\rho}{2\rho\sigma^2} \sum_{i=1}^{\ell} (\mu_i - \theta)^2 \right\}.$$

The CPDs for θ and σ^2 are normal and inverse gamma densities respectively, but the CPD of ρ is not simple and we can use the grid method to sample it.

We have used a more efficient Gibbs sampler, which is obtained by using the following decomposition,

$$\pi(\boldsymbol{\mu},\mathbf{r},\mathbf{t},\epsilon,\theta,\sigma^2,\rho\,|\,\mathbf{y}) = \pi_1(\boldsymbol{\mu},\theta,\sigma^2,\rho\,|\,\mathbf{r},\mathbf{t},\epsilon,\mathbf{y})\,\pi_2(\mathbf{r},\mathbf{t},\epsilon\,|\,\mathbf{y}),$$

where $(\boldsymbol{\mu},\theta,\sigma^2,\rho)$ can be sampled simultaneously from $\pi_1(\boldsymbol{\mu},\theta,\sigma^2,\rho\,|\,\mathbf{r},\mathbf{t},\epsilon,\mathbf{y})$ via the multiplication rule of probability; see Nandram, Toto and Choi (2011). This is a blocked Gibbs sampler, which is more efficient than a systematic-scan Gibbs sampler (e.g., see Tan and Hobert, 2009), and fewer number of runs of the Gibbs sampler are needed.

Remarks

a. We can prove that the joint posterior density is proper. We can integrate out the $\mu_i,\theta,\sigma^2,\epsilon,$ and we will be left with $\pi(\mathbf{r},\mathbf{t},\rho\,|\,\mathbf{y})$ if we assume $\ell \geq 2$. We can then transform $t_i = \frac{\gamma_i}{1-\gamma_i}$ (i.e., $0 < \gamma_i < 1$) and with the assumption that γ_i is strictly bounded away from 0 and 1, and similarly for $\rho,$ we can prove the joint posterior density is proper.

b. We will need the masked total of the sensitive counties to be exactly the same as the observed total. This is done by benchmarking (raking) in the post-processing; see earlier comments about the sensitivity of the target. Post-processing maintains the same level of privacy protections as the noisy measurements (Dwork and Roth, 2014, Proposition 2.1). The observed state total in TAB is not used in the masking procedure (model fitting), and we can think of them as external data (Williams and Bowen, 2023), and the actual values of the farm level acres are not used in the raking (benchmarking) procedure.

c. Benchmarking is done on the original scale as follows. Let $\tilde{T}_1,\ldots,\tilde{T}_\ell$ denote the predicted values of the county totals, and B denote the total of the observed confidential data. We want $T_1,\ldots,T_\ell$ such $\sum_{i=1}^{\ell} T_i = B,$ which are

$$T_i = \left(\frac{B}{\sum_{i=1}^{\ell}\tilde{T}_i}\right)\tilde{T}_i,\ i = 1,\ldots,\ell.$$

Note that $\sum_{i=1}^{\ell} T_i = B,$ and in this post-processing stage, the observed confidential data are not used. This is done in the output analysis for each of the "good iterates" from the Gibbs sampler to obtain the empirical distributions of the masked data. This raking procedure (proportional allocation) is standard in survey sampling, but it can be improved using a model-based analysis.

3.2 Robust mixture model

For the study variable, we can use a mixture model with stick-breaking weights. This robust mixture model will provide better robustness against non-normality and large acres (outliers), and in this sense, better than the parametric model. This is within the same spirit of the Dirichlet process.

Let $n=\sum_{i=1}^{\ell} n_i$, the mixture model places the $y_{ij}, j=1,\ldots,n_i, i=1,\ldots,\ell$ into clusters. Let $n=\sum_{i=1}^{\ell} n_i$ and let the number of clusters of the y_{ij} be n_o, where $n_o \le n$. Then, we assume independence of the y_{ij} with

$$p(y_{ij}, d_{ij} \mid n_o) = \prod_{s=1}^{n_o} \left[p_s \text{ Normal}\left\{ \mu_i + (2r_i - 1)\, t_i + \eta_s, \sigma^2 \right\} \right]^{I(di_j = s)}, n_o \le n,$$

where the d_{ij} are latent variables with $d_{ij}=s, j=1,\ldots,n_i, i=1,\ldots,\ell,$ means that the j^{th} farm within the i^{th} county is mapped into the s^{th} cluster. Also, p_s are stick-breaking weights (e.g., Ishwaran and James, 2001),

$$p_1 = \nu_1,\ p_2 = \nu_2(1-\nu_1), \ldots,\ p_s = \prod_{k=1}^{s-1}(1-\nu_k),\ \sum_{k=1}^{s} p_k = 1,$$

$$\nu_s \overset{\text{ind}}{\sim} \text{Beta}\left\{ \delta_1^s \frac{\delta_2}{1-\delta_2}, (1-\delta_1^s)\frac{\delta_2}{1-\delta_2} \right\}, 0 < \delta_1 < \frac{1}{2} < \delta_2 < 1.$$

That is, the $(p_1,\ldots,p_s)$ have a generalized Dirichlet distribution. The distributions of the ν_s, specified here, are like those in the two-parameter Pitman-Yor process. The stick-breaking process comes up because from a stick of unit length, $p_1=\nu_1$ is broken off with remainder $1-\nu_1$; and then ν_2 of the remainder is broken off to get $p_2=\nu_2(1-\nu_1)$, and so on. Note that $\sum_{s=1}^{n_o} p_s = 1$. The Dirichlet process is another stick-breaking process (Sethuraman, 1994; Ishwaran and James, 2001), but not as general as the one described here.

Then, the joint density of $(\mathbf{y},\mathbf{d})$ is

$$p(\mathbf{y},\mathbf{d} \mid n_o) = \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \prod_{s=1}^{n_o} \left[p_s \text{ Normal}\left\{ \mu_i + (2r_i - 1)\, t_i + \eta_s, \sigma^2 \right\} \right]^{I(d_{ij} = s)},$$

and a priori for clustering, we assume

$$\eta_s \overset{\text{ind}}{\sim} \text{Normal}\left\{ 0, \frac{\rho_1}{1-\rho_1}\sigma^2 \right\}, s = 1,\ldots,n_o.$$

Unlike the parametric model, this mixtue model will accommodate the outliers (large acreages) by probabilistically partitioning the data into clusters; it is also robust against non-normality. Note that we are assuming that the $y_{ij} - \mu_i$, not the y_{ij}, have the standard differential privacy, since $\text{E}(y_{ij} - \mu_i \mid n_o) = (2r_i - 1)\, t_i, j = 1,\ldots,n_i, i = 1,\ldots,\ell,$ the Laplace differential privacy mechanism.

The rest of the model (i.e., the assumptions below) is similar to the parametric model. That is,

$$\mu_i \overset{\text{ind}}{\sim} \text{Normal}\left\{ \theta, \frac{\rho_2}{1-\rho_2}\sigma^2 \right\},\ r_i \overset{\text{ind}}{\sim} \text{Bernoulli}\left\{ \frac{1}{2} \right\},\ t_i \mid \epsilon \overset{\text{ind}}{\sim} \text{Gamma}\left\{ 1, a_{oi}\epsilon \right\}, i = 1,\ldots,\ell,$$

$$\pi(\theta, \sigma^2, \epsilon, \rho_1, \rho_2) \propto \frac{1}{\sigma^2} e^{-\epsilon}, a_o < \epsilon < b_o$$

(e.g., $a_o = 0, b_o = 2.25$, same as in the parametric model). We can fit the complete model using the blocked Gibbs sampler; e.g., there is one block on $\mathbf{d}$ and another block on $(\mathbf{v}, \delta_1, \delta_2)$. By starting with $n_o = 15$, say, we can make a preliminary set of clusters by partitioning the data, and a partition matrix, E, an $n \times n_o$ incidence matrix, can be obtained. This will give us $(\boldsymbol{\mu}, \mathbf{r}, \mathbf{t}, \boldsymbol{\eta}, \sigma^2)$, and the stick-breaking weights can be obtained. Let $\mathbf{e}'_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, \ell,$ denote the $(ij)^{\text{th}}$ row of E. Then,

$$y_{ij} \mid d_{ij} \overset{\text{ind}}{\sim} \text{Normal}\left\{\mu_i + (2r_i - 1)\, t_i + \mathbf{e}'_{ij}\boldsymbol{\eta}, \sigma^2\right\}, j = 1, \ldots, n_i, i = 1, \ldots, \ell.$$

It is straightforward to write down the rest of the CPDs to run the Gibbs sampler.

However, we note that the CPD of all parameters (given $\mathbf{d}$) is

$$\begin{aligned}\pi(\boldsymbol{\mu}, \mathbf{r}, \mathbf{t}, \boldsymbol{\eta}, \epsilon, \theta, \rho_1, \rho_2, \sigma^2 \mid \mathbf{d}, \mathbf{y}) \propto & \left(\frac{1}{\sigma^2}\right)^{(n+n_o+\ell)/2+1} \left(\frac{1-\rho_1}{\rho_1}\right)^{n_o/2} \left(\frac{1-\rho_2}{\rho_2}\right)^{\ell/2} e^{-\epsilon} I_{[a_o, b_o]}(\epsilon) \\ & \times \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}(y_{ij} - \mu_i - (2r_i - 1)\, t_i - \mathbf{e}'_{ij}\boldsymbol{\eta})^2\right\} \times \exp\left\{-\frac{1-\rho_1}{2\rho_1\sigma^2}\sum_{s=1}^{n_o}\eta_s^2\right\} \\ & \times \exp\left\{-\frac{1-\rho_2}{2\rho_2\sigma^2}\sum_{i=1}^{\ell}(\mu_i - \theta)^2\right\} \times \prod_{i=1}^{\ell}(a_{oi}\epsilon) \times \exp\{-a_{oi}\epsilon t_i\}.\end{aligned}$$

3.3 Over-shrinkage and prediction

We describe a simple procedure to overcome over-shrinkage (an on-going problem in small area estimation) in both models. There are other ways to do so (e.g., global-local pooling), but this complication is not necessary. This is an extension of the two models. We also show how to do the prediction in both models with an adjustment to increase utility using the absolute relative error (ARE).

To help mitigate over-shrinkage and avoid non-identifiability issues, we now add a common feature to both models. That is, we have partitioned the counties using the weighted average of the farm acreages into deciles. Specifically, we look at

$$\overline{y}_i = \frac{\sum_{j=1}^{n_i} w_{ij} y_{ij}}{\sum_{j=1}^{n_i} w_{ij}}, \; i = 1, \ldots, \ell,$$

where w_{ij} are the Census weights, and we partition the $\overline{y}_i$ into deciles (ten intervals). Let q_0 and q_{10} denote the minimum and maximum values of the $\overline{y}_i$, and $q_1, \ldots, q_9$ the $10^{\text{th}}, \ldots, 90^{\text{th}}$ percentiles of the $\overline{y}_i$. So the the deciles are $[q_0, q_1), [q_1, q_2), \ldots, [q_9, q_{10}]$. Let $C_s, s = 1, \ldots, 10,$ denote the set of counties in these deciles. For $s = 1, \ldots, 10,$ denote the cardinality of these C_s by g_s, which is approximately 8.

Then, we assume that

$$\mu_i \overset{\text{ind}}{\sim} \text{Normal}\left(\theta_s, \frac{\rho}{1-\rho}\sigma^2\right), \; i \in C_s, s = 1, \ldots, 10.$$

In the original parametric model, there is a single θ. This is coupled with our standard assumptions in the parametric model and robust mixture model. For example, in the parametric model,

$$y_{ij} \overset{\text{ind}}{\sim} \text{Normal}\{\mu_i + (2r_i - 1)\, t_i, \sigma^2\},\ j = 1,\ldots,n_i, i = 1,\ldots,\ell,$$

with the same assumptions on $r_i, t_i,$ and ϵ in the original parametric model. We also have

$$\pi(\boldsymbol{\theta}) = 1,$$

where the θ_s are ordered as $-\infty < \theta_1 < \ldots < \theta_{10} < \infty$. A similar adjustment is made in the robust mixture model.

Bayesian predictive inference is used to provide masked county estimates. These estimates are then benchmarked as discussed earlier. We use the population model to do the prediction; the sample model with the added Laplace noise already provides the adjustment to the parameters of the population model.

For the parametric model, prediction is done using the population model,

$$Y_{ij} \mid \mu_i, \sigma^2 \sim \text{Normal}(\mu_i, \sigma^2),\ j = 1,\ldots,n_i, i = 1,\ldots,\ell,$$

and this is subjected to any transformation used. The county estimates are then

$$Y_i = \sum_{j=1}^{n_i} Y_{ij}, i = 1,\ldots,\ell,$$

and the posterior predictive distributions of the masked county acreages are

$$\pi(Y_i \mid \mathbf{y}) = \int f(Y_i \mid \mu_i, \sigma^2, \mathbf{y})\, \pi(\mu_i, \sigma^2 \mid \mathbf{y})\, d\mu_i\, d\sigma^2 = \int f(Y_i \mid \mu_i, \sigma^2)\, \pi(\mu_i, \sigma^2 \mid \mathbf{y})\, d\mu\, d\sigma^2,$$

where the posterior density, $\pi(\mu_i, \sigma^2 \mid \mathbf{y}),$ comes from the sample model, and

$$f(Y_i \mid \mu_i) \overset{\text{ind}}{\sim} \text{Normal}\left(n_i\mu_i, n_i\sigma^2\right), i = 1,\ldots,\ell.$$

Of course, this is done in the post-processing of the Gibbs sampler coupled with the benchmarking as discussed earlier. With a transformation, we need to predict the Y_{ij} individually. That is, $Y_{ij} \mid \mu_i \sim$ Normal $(\mu_i, \sigma^2),\ j = 1,\ldots,n_i, i = 1,\ldots,\ell,$ and then back transformation is done.

Like the parametric model, prediction is done in the population model (robust mixture),

$$y_{ij} \mid \mu_i, \sigma^2, \boldsymbol{\eta}, \mathbf{p}, n_o \overset{\text{ind}}{\sim} \sum_{s=1}^{n_o} p_s\, \text{Normal}\left(\mu_i + \eta_s, \sigma^2\right), j = 1,\ldots,n_i, i = 1,\ldots,\ell.$$

This population model is fed by the sample model, and data masking is done in exactly the same manner as in the parametric model.

We can increase utility by putting a bound on the AREs. On the transformed scale (i.e., square root) for each farm, we can assume that

$$\left| \frac{P^2 - O^2}{O^2} \right| < a, 0 < a < 1,$$

where P is the predicted acreage and O is the observed acreage on the transformed scale. Then, $O\sqrt{1-a} < P < O\sqrt{1+a}$. For example, if we choose $a = 1$, $a = 0.75$ or $a = 0.5$, we have $0 < P < O\sqrt{2}$, $O\sqrt{0.25} < P < O\sqrt{1.75}$ or $O\sqrt{0.5} < P < O\sqrt{1.5}$. In each example, this is a mild violation of the post-processing theorem that we can tolerate to increase utility a bit. To be conservative, we have used $a = 0.75$; $a = 1$ leads to AREs bigger than unity and $a = 0.50$ leads to AREs that are artificially low on the original untransformed scale. This adjustment is also necessary to prevent too small (or large) masked county totals relative to the original confidential observations.

4. Comparisons of the two models

We compare the two models using simulated data on harvested acres for COM in TAB of STAT; STAT has COM grown in the 75 STAT counties and the weighted number (N) of farms in a county varies from 2 to 97. [In both models, we have kept ϵ in (0.000, 2.250).] Recall that there are 17 suppressed counties, and 58 non-suppressed counties partitioned into 38 non-sensitive counties and 20 sensitive counties. As stated earlier, the acres for the non-sensitive counties remain unmasked from the original observed acres of the Census, and the weighted acres are reported for these counties. The suppressed counties are never reported. The models are fit to the 58 non-suppressed counties. We also compare these two small area models with an individual area (county) model. In our tables and graphs, we report results for only the 20 sensitive counties.

The Gibbs sampler was used for fitting both models using a long run. Monitoring of the Gibbs samplers is obtained using the Geweke tests (Geweke, 1992) and the effective sample sizes for the hyperparameters to assess convergence and strong mixing. Again, post-processing, where we benchmarked the masked data to the observed state total, maintains the same level of privacy protections as the noisy measurements (Dwork and Roth, 2014, Proposition 2.1) is counted in the computation time. All computations were performed using WPI's High Performance Computer Cluster (Turing). [The cluster features 3,492 CPUs, ranging from Intel E5-2,695 v4 processors to Intel Scalable Gold 6,248 and AMO Epyc 7,543 processors, and 40.5TB of memory. The cluster has GPU-accelerated segments with 60 GPUs, including 20x 80GB A100's and 2x H100's.]

For posterior summaries, we have used the posterior mean (PM), posterior standard deviation (PSD), numerical standard error (NSE), posterior coefficient of variation (PCV) and 95% highest posterior density interval (HPDI; unimodality is assumed). We have used the posterior coefficient of variation (PCV) as a measure of reliability of the Bayesian procedure. If PCV $\leq$ 0.30, reliability is high (H); if 0.30 < PCV $\leq$ 0.75, reliability is medium (M); and if PCV > 0.75, reliability is low (L). On the original scale, for each county, again we define utility as the ARE,

$$\text{ARE} = \left| (\text{PM} - \text{Obs}) / \text{Obs} \right|,$$

where Obs refers to the observed data, with small values being good utility. To show different degrees of utility (excellent to poor), we have looked at the AREs to get E: ARE ≤ 0.05 (excellent); V: $0.05 \leq$ ARE < 0.10 (very good); G: $0.10 \leq$ ARE < 0.25 (good); P: ARE ≥ 0.25 (poor). (We may call the utility for individual counties local utility.)

In addition, we have considered two global measures of utility. First, we have used the propensity score mean-squared error (pMSE) as described in Snoke, Raab, Nowok, Dibben and Slavkovic (2018); values of pMSE close to zero give high global utility. This is a good measure of global utility; other measures are presented by Snoke et al. (2018) such as the pMSE-ratio, the standardized pMSE, but these are not appropriate because they are based on simple random sampling. The propensity scores are calculated using logistic regression with the response variable (quadratic regression) as the covariate (nonignorability); other covariates are not used in this analysis. When we add the square of the response variable, we get a small improvement. Second, we have used the Wilcoxon signed rank test (WSRT), where we study how close the posterior mean (or each "good" iterate from the Gibbs sampler) is to the observed acres (matched pairs). The WSRT gives an overall assessment of utility with a large p-value meaning high utility and a small p-value meaning lower utility. We prefer to use the p-values because they are calibrated, but the test statistics are not. [A reviewer correctly pointed out that the Kolmogorov-Smirnov (KS) test is not appropriate because the two samples are highly correlated.]

For the parametric model, we ran the Gibbs sampler 300,000 times, used 50,000 iterates as a burn-in, and took a systematic sample of every 250 runs. This took roughly 2 minutes on Turing. The Geweke tests and the effective sample sizes show good performance of the Gibbs sampler for $\boldsymbol{\theta}, \sigma^2, \rho, \epsilon;$ the p-value of the Geweke test and effective sample size for ϵ are 0.937 and 1,000.

For the robust mixture model, because it is much more complex than the parametric model, one would expect more runs are needed for proper convergence, but this is not true. In fact, we ran the Gibbs sampler 200,000 times, used 50,000 iterates as a burn-in, took a systematic sample of every 150 runs, and convergence (Geweke test and effective sample sizes) was monitored. This took about 13 minutes on Turing. Again, the Geweke tests and the effective sample sizes show good performance of the Gibbs sampler for $\boldsymbol{\theta}, \sigma^2, \rho_1, \rho_2, \delta_1, \delta_2, n_o, \epsilon;$ the p-value of the Geweke test and effective sample size for ϵ are 0.902 and 1,000.

The mixture model is very flexible, and it adds credibility to our Bayesian procedure because it can accommodate extreme farm acres. This can be seen by looking at the number of components (i.e., clusters), $n_o,$ an important parameter in the robust mixture model. The p-value of the Geweke test and effective sample size for n_o are 0.587 and 1,000. For $n_o,$ PM = 5.227, PSD = 1.383, NSE = 0.038, PCV = 0.264 and the 95% HPDI is (3.00, 8.00); posterior inference for the number of clusters is reliable as measured by the PCV. Therefore, it is clear that the robust mixture model can separate the county acres into different groups (ordinary data, mild outliers, severe outliers, etc.) and clearly the parametric model cannot do this. However, both models use many runs for convergence, with the parametric model requiring more runs, but the computation to convergence for the parametric model took much less time than the robust mixture model.

We have computed the local (approximately global) sensitivity for the masked data (here county posterior means). Under the parametric model, we got 3,553 (5.967 standard units) for all counties and 585 (6.797 standard units) for sensitive counties. Under the robust mixture model, we got 3,492 (5.660 standard units) for all counties and 583 (6.623 standard units) for sensitive counties. These are similar, and as expected, they are similar to those of the original observed confidential data.

For comparison, we consider an individual area (county) model. Let $y_j, j = 1,\ldots,n,$ denote the farm-level weighted acreages, where n is the number of farms in this county. Note that we are only considering sensitive counties, and we use a version of the small area parametric model for a single county. Therefore, we assume

$$y_j \mid \mu, r, t, \sigma^2 \overset{\text{ind}}{\sim} \text{Normal}\,(\mu + (2r-1)\,t, \sigma^2),\ j = 1,\ldots,n.$$

Let t_o denote the scale (same as in the two small area models) for this county only. For the differential privacy mechanism, we assume

$$r \sim \text{Bernoulli}\left(\frac{1}{2}\right), t \sim \text{Gamma}\left(1, \frac{\epsilon}{t_o}\right), \epsilon \sim \text{Gamma}\,(1,1),\ a_o < \epsilon < b_o,$$

where again, we take $a_o = 0.0$ and $b_o = 2.25$, and a priori, $\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$. Using Bayes' theorem, the joint posterior density is

$$\pi(\mu, \sigma^2, r, t, \epsilon \mid \mathbf{y}) \propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{1}{2\sigma^2}\{(n-1)\,s^2 + n(\bar{y} - (2r-1)\,t - \mu)^2\}\right\} \frac{\epsilon}{t_o} e^{-\frac{\epsilon}{t_o}t} e^{-\epsilon} I_{(0,b_o)}(\epsilon), t > 0.$$

This posterior density can be fit using the Gibbs sampler, but one needs to be careful because $(r, t, \mu, \sigma^2, \epsilon)$ are not fully identifiable. To help mitigate this non-identifiability, we use the two blocks, (μ, σ^2) and $(r, t, \epsilon),$ to run the Gibbs sampler.

The Gibbs sampler is run 65,000 times, a burn-in of 15,000 is used, and a systematic sample of every fiftieth one is selected to get a sample of size, 1,000. All 20 Gibbs samplers took just about 11 minutes on Turing. The Geweke p-values and the effective sample sizes are mostly acceptable when they are run for $(\mu, \sigma^2, t, \epsilon)$; for ϵ the p-values are all large enough and the effective sample sizes are mostly 1,000. Benchmarking for the 20 counties is done as in the two small area models. In Tables 4.1 to 4.3 and Figures 4.1 to 4.4, we compare the three models using the AREs, PCVs, PMs, PSDs. We include only the 20 sensitive counties in these tables and figures, because suppressed counties are never presented and the weighted acres of non-sensitive counties are presented without masking.

In Tables 4.1 to 4.3, we can see many things. In Table 4.1, the mixture model is generally better than the parametric model when we look at the AREs and the PCVs; the PSDs are also smaller under the mixture model. In Table 4.2, the parametric model is better than the individual area model, and in Table 4.3, the mixture model is better than the individual area model, when we look at the AREs and the PCVs; the PSDs are also smaller under the mixture model. Therefore, the mixture model is much more reliable than the two competitors, especially the individual area model.

Table 4.1
Comparison of the parametric model and the mixture model using posterior summaries about masked data for sensitive counties

C	*N*	Obs	PM	ARE	Utility	PSD	NSE	PCV	Rel	95% HPDI
a. Parametric Model										
8	14	95.400	107.505	0.127	G	25.169	0.664	0.234	H	(62.272, 155.343)
12	7	58.000	61.849	0.066	V	25.518	0.914	0.413	M	(26.282, 112.268)
15	5	60.500	63.825	0.055	V	24.326	0.831	0.381	M	(27.497, 110.899)
17	5	44.800	50.455	0.126	G	14.785	0.603	0.293	H	(23.225, 76.857)
18	7	89.300	94.899	0.063	V	33.710	1.066	0.355	M	(41.524, 163.502)
19	10	25.500	29.714	0.165	G	7.679	0.198	0.258	H	(15.437, 45.922)
28	12	84.300	92.165	0.093	V	27.763	0.726	0.301	M	(45.029, 144.265)
31	24	178.000	165.012	0.073	V	50.435	1.452	0.306	M	(82.448, 271.373)
43	59	518.000	525.721	0.015	E	79.725	2.613	0.152	H	(383.559, 689.560)
46	5	102.500	98.845	0.036	E	41.108	1.206	0.416	M	(43.179, 186.845)
48	5	60.500	67.643	0.118	G	20.298	0.606	0.300	M	(34.336, 108.325)
49	16	121.400	133.035	0.096	V	33.157	1.205	0.249	H	(73.781, 198.713)
50	7	25.600	29.343	0.146	G	10.556	0.370	0.360	M	(11.478, 48.428)
57	11	218.400	193.913	0.112	G	56.509	1.646	0.291	H	(116.698, 319.623)
59	5	41.300	46.875	0.135	G	15.068	0.478	0.321	M	(21.675, 75.556)
61	9	50.500	57.227	0.133	G	16.414	0.519	0.287	H	(26.724, 89.722)
62	29	455.300	439.421	0.035	E	84.016	2.731	0.191	H	(276.284, 596.880)
70	6	110.600	113.344	0.025	E	37.461	1.292	0.331	M	(51.437, 188.677)
72	15	58.000	66.782	0.151	G	17.617	0.570	0.264	H	(37.101, 100.482)
76	31	417.800	378.128	0.095	V	75.786	3.027	0.200	H	(251.244, 539.671)
b. Mixture Model										
8	14	95.400	106.886	0.120	G	23.110	0.745	0.216	H	(62.678, 150.491)
12	7	58.000	60.972	0.051	V	22.601	0.710	0.371	M	(24.259, 102.284)
15	5	60.500	65.248	0.078	V	22.653	0.671	0.347	M	(27.306, 105.988)
17	5	44.800	49.520	0.105	G	13.503	0.467	0.273	H	(26.225, 76.459)
18	7	89.300	93.059	0.042	E	29.809	0.890	0.320	M	(41.293, 149.528)
19	10	25.500	28.793	0.129	G	6.846	0.200	0.238	H	(15.460, 41.969)
28	12	84.300	91.069	0.080	V	24.121	0.814	0.265	H	(45.356, 136.532)
31	24	178.000	171.431	0.037	E	49.926	1.230	0.291	H	(95.786, 277.416)
43	59	518.000	519.965	0.004	E	70.480	2.315	0.136	H	(384.103, 654.173)
46	5	102.500	98.059	0.043	E	41.151	1.429	0.420	M	(41.235, 177.710)
48	5	60.500	66.654	0.102	G	17.569	0.596	0.264	H	(36.033, 102.083)
49	16	121.400	131.711	0.085	V	30.520	1.000	0.232	H	(73.449, 188.965)
50	7	25.600	28.741	0.123	G	9.433	0.241	0.328	M	(11.420, 45.431)
57	11	218.400	201.056	0.079	V	57.575	2.009	0.286	H	(116.466, 326.679)
59	5	41.300	45.305	0.097	V	13.797	0.496	0.305	M	(20.169, 70.867)
61	9	50.500	56.933	0.127	G	15.100	0.497	0.265	H	(27.731, 84.088)
62	29	455.300	443.668	0.026	E	73.115	2.319	0.165	H	(309.219, 592.471)
70	6	110.600	111.124	0.005	E	35.230	1.212	0.317	M	(53.610, 177.132)
72	15	58.000	63.627	0.097	V	15.276	0.428	0.240	H	(37.934, 94.967)
76	31	417.800	381.878	0.086	V	70.250	1.582	0.184	H	(247.661, 516.522)

Note: Absolute relative error, ARE = |(PM-OBS)/OBS|. We use for Utility (E: excellent, V: very good, G: good, P: poor); Rel (Reliability), (H: high, M: medium, L: low). Absolute relative error (ARE); highest posterior density interval (HPDI); numerical standard error (NSE); Observed data (Obs); posterior coefficient of variation (PCV); posterior mean (PM); posterior standard deviation (PSD).

Table 4.2
Comparison of the parametric model and the individual model using posterior summaries about masked data for sensitive counties

C	*N*	Obs	PM	ARE	Utility	PSD	NSE	PCV	Rel	95% HPDI
a. Parametric Model										
8	14	95.400	107.505	0.127	G	25.169	0.664	0.234	H	(62.272, 155.343)
12	7	58.000	61.849	0.066	V	25.518	0.914	0.413	M	(26.282, 112.268)
15	5	60.500	63.825	0.055	V	24.326	0.831	0.381	M	(27.497, 110.899)
17	5	44.800	50.455	0.126	G	14.785	0.603	0.293	H	(23.225, 76.857)
18	7	89.300	94.899	0.063	V	33.710	1.066	0.355	M	(41.524, 163.502)
19	10	25.500	29.714	0.165	G	7.679	0.198	0.258	H	(15.437, 45.922)
28	12	84.300	92.165	0.093	V	27.763	0.726	0.301	M	(45.029, 144.265)
31	24	178.000	165.012	0.073	V	50.435	1.452	0.306	M	(82.448, 271.373)
43	59	518.000	525.721	0.015	E	79.725	2.613	0.152	H	(383.559, 689.560)
46	5	102.500	98.845	0.036	E	41.108	1.206	0.416	M	(43.179, 186.845)
48	5	60.500	67.643	0.118	G	20.298	0.606	0.300	M	(34.336, 108.325)
49	16	121.400	133.035	0.096	V	33.157	1.205	0.249	H	(73.781, 198.713)
50	7	25.600	29.343	0.146	G	10.556	0.370	0.360	M	(11.478, 48.428)
57	11	218.400	193.913	0.112	G	56.509	1.646	0.291	H	(116.698, 319.623)
59	5	41.300	46.875	0.135	G	15.068	0.478	0.321	M	(21.675, 75.556)
61	9	50.500	57.227	0.133	G	16.414	0.519	0.287	H	(26.724, 89.722)
62	29	455.300	439.421	0.035	E	84.016	2.731	0.191	H	(276.284, 596.880)
70	6	110.600	113.344	0.025	E	37.461	1.292	0.331	M	(51.437, 188.677)
72	15	58.000	66.782	0.151	G	17.617	0.570	0.264	H	(37.101, 100.482)
76	31	417.800	378.128	0.095	V	75.786	3.027	0.200	H	(251.244, 539.671)
b. Individual Model										
8	14	95.400	79.194	0.170	G	37.097	1.764	0.468	M	(22.099, 147.552)
12	7	58.000	37.640	0.351	P	16.299	0.697	0.433	M	(16.472, 74.560)
15	5	60.500	43.954	0.273	P	18.193	0.617	0.414	M	(16.337, 80.381)
17	5	44.800	40.429	0.098	V	22.084	1.045	0.546	M	(10.400, 76.628)
18	7	89.300	62.966	0.295	P	21.326	0.883	0.339	M	(27.800, 103.893)
19	10	25.500	22.946	0.100	G	11.066	0.571	0.482	M	(5.438, 41.253)
28	12	84.300	64.234	0.238	G	26.314	1.085	0.410	M	(21.397, 115.674)
31	24	178.000	130.865	0.265	P	62.570	2.948	0.478	M	(41.944, 255.148)
43	59	518.000	351.244	0.322	P	114.566	6.006	0.326	M	(170.364, 587.641)
46	5	102.500	69.292	0.324	P	29.474	1.136	0.425	M	(28.365, 131.002)
48	5	60.500	49.534	0.181	G	17.686	0.568	0.357	M	(16.756, 81.967)
49	16	121.400	93.514	0.230	G	41.769	1.852	0.447	M	(31.586, 177.478)
50	7	25.600	22.040	0.139	G	12.263	0.435	0.556	M	(5.788, 42.880)
57	11	218.400	142.618	0.347	P	54.434	1.718	0.382	M	(61.851, 267.875)
59	5	41.300	30.302	0.266	P	11.468	0.431	0.378	M	(12.711, 54.605)
61	9	50.500	41.648	0.175	G	18.513	0.648	0.445	M	(10.245, 74.589)
62	29	455.300	319.743	0.298	P	110.496	4.874	0.346	M	(151.884, 550.197)
70	6	110.600	78.447	0.291	P	26.990	1.073	0.344	M	(35.761, 133.664)
72	15	58.000	43.728	0.246	G	16.018	0.659	0.366	M	(16.028, 76.549)
76	31	417.800	280.711	0.328	P	101.139	6.141	0.360	M	(125.131, 500.439)

Note: Absolute relative error, ARE = |(PM-OBS)/OBS|. We use for Utility (E: excellent, V: very good, G: good, P: poor); Rel (Reliability), (H: high, M: medium, L: low). Absolute relative error (ARE); highest posterior density interval (HPDI); numerical standard error (NSE); Observed data (Obs); posterior coefficient of variation (PCV); posterior mean (PM); posterior standard deviation (PSD).

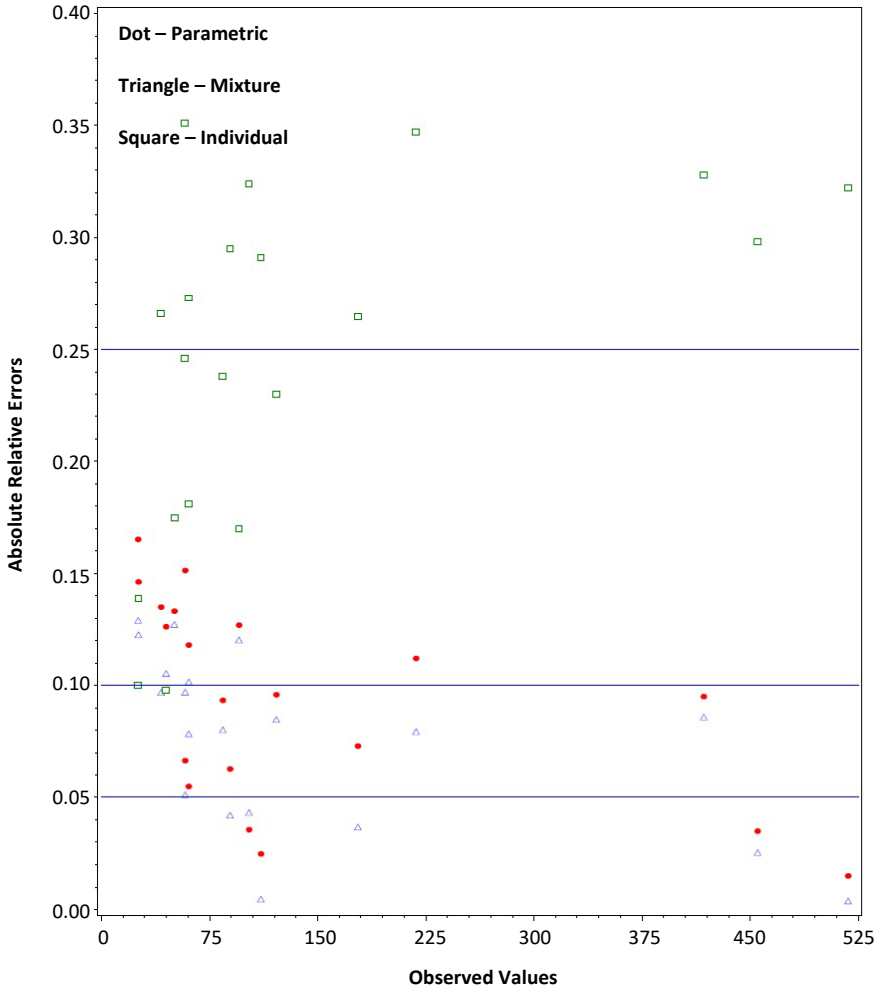
Table 4.3
Comparison of the mixture model and the individual model using posterior summaries about masked data by sensitive county

C	*N*	Obs	PM	ARE	Utility	PSD	NSE	PCV	Rel	95% HPDI
a. Mixture Model										
8	14	95.400	106.886	0.120	G	23.110	0.745	0.216	H	(62.678, 150.491)
12	7	58.000	60.972	0.051	V	22.601	0.710	0.371	M	(24.259, 102.284)
15	5	60.500	65.248	0.078	V	22.653	0.671	0.347	M	(27.306, 105.988)
17	5	44.800	49.520	0.105	G	13.503	0.467	0.273	H	(26.225, 76.459)
18	7	89.300	93.059	0.042	E	29.809	0.890	0.320	M	(41.293, 149.528)
19	10	25.500	28.793	0.129	G	6.846	0.200	0.238	H	(15.460, 41.969)
28	12	84.300	91.069	0.080	V	24.121	0.814	0.265	H	(45.356, 136.532)
31	24	178.000	171.431	0.037	E	49.926	1.230	0.291	H	(95.786, 277.416)
43	59	518.000	519.965	0.004	E	70.480	2.315	0.136	H	(384.103, 654.173)
46	5	102.500	98.059	0.043	E	41.151	1.429	0.420	M	(41.235, 177.710)
48	5	60.500	66.654	0.102	G	17.569	0.596	0.264	H	(36.033, 102.083)
49	16	121.400	131.711	0.085	V	30.520	1.000	0.232	H	(73.449, 188.965)
50	7	25.600	28.741	0.123	G	9.433	0.241	0.328	M	(11.420, 45.431)
57	11	218.400	201.056	0.079	V	57.575	2.009	0.286	H	(116.466, 326.679)
59	5	41.300	45.305	0.097	V	13.797	0.496	0.305	M	(20.169, 70.867)
61	9	50.500	56.933	0.127	G	15.100	0.497	0.265	H	(27.731, 84.088)
62	29	455.300	443.668	0.026	E	73.115	2.319	0.165	H	(309.219, 592.471)
70	6	110.600	111.124	0.005	E	35.230	1.212	0.317	M	(53.610, 177.132)
72	15	58.000	63.627	0.097	V	15.276	0.428	0.240	H	(37.934, 94.967)
76	31	417.800	381.878	0.086	V	70.250	1.582	0.184	H	(247.661, 516.522)
b. Individual Model										
8	14	95.400	79.194	0.170	G	37.097	1.764	0.468	M	(22.099, 147.552)
12	7	58.000	37.640	0.351	P	16.299	0.697	0.433	M	(16.472, 74.560)
15	5	60.500	43.954	0.273	P	18.193	0.617	0.414	M	(16.337, 80.381)
17	5	44.800	40.429	0.098	V	22.084	1.045	0.546	M	(10.400, 76.628)
18	7	89.300	62.966	0.295	P	21.326	0.883	0.339	M	(27.800, 103.893)
19	10	25.500	22.946	0.100	G	11.066	0.571	0.482	M	(5.438, 41.253)
28	12	84.300	64.234	0.238	G	26.314	1.085	0.410	M	(21.397, 115.674)
31	24	178.000	130.865	0.265	P	62.570	2.948	0.478	M	(41.944, 255.148)
43	59	518.000	351.244	0.322	P	114.566	6.006	0.326	M	(170.364, 587.641)
46	5	102.500	69.292	0.324	P	29.474	1.136	0.425	M	(28.365, 131.002)
48	5	60.500	49.534	0.181	G	17.686	0.568	0.357	M	(16.756, 81.967)
49	16	121.400	93.514	0.230	G	41.769	1.852	0.447	M	(31.586, 177.478)
50	7	25.600	22.040	0.139	G	12.263	0.435	0.556	M	(5.788, 42.880)
57	11	218.400	142.618	0.347	P	54.434	1.718	0.382	M	(61.851, 267.875)
59	5	41.300	30.302	0.266	P	11.468	0.431	0.378	M	(12.711, 54.605)
61	9	50.500	41.648	0.175	G	18.513	0.648	0.445	M	(10.245, 74.589)
62	29	455.300	319.743	0.298	P	110.496	4.874	0.346	M	(151.884, 550.197)
70	6	110.600	78.447	0.291	P	26.990	1.073	0.344	M	(35.761, 133.664)
72	15	58.000	43.728	0.246	G	16.018	0.659	0.366	M	(16.028, 76.549)
76	31	417.800	280.711	0.328	P	101.139	6.141	0.360	M	(125.131, 500.439)

Note: Absolute relative error, ARE = |(PM-OBS)/OBS|. We use for Utility (E: excellent, V : very good, G: good, P: poor); Rel (Reliability) (H: high, M: medium, L: low). Absolute relative error (ARE); highest posterior density interval (HPDI); numerical standard error (NSE); Observed data (Obs); posterior coefficient of variation (PCV); posterior mean (PM); posterior standard deviation (PSD).

In Figure 4.1, we compare the three models using the AREs of the 20 sensitive counties. The horizontal straight lines are at ARE = 0.05, 0.10, 0.25. Both the parametric model and the mixture model have AREs below the 0.05 line, but none from the individual area model with most of its points lying above the 0.25 line. For the mixture model, the points are generally lower than the parametric model.

Figure 4.1 Plots of the county absolute relative errors (AREs) versus the observed values by model



In Figure 4.2, we compare the three models using the PCVs of the 20 sensitive counties. The horizontal straight lines are at PCV = 0.30 and PCV = 0.75. It is good that the PCVs under the robust mixture model are mostly smaller than those under the parametric model, and therefore under this model the robust mixture model is a bit more reliable. All the points under individual area model are above the 0.30 line, and therefore the individual area model is very unreliable.

In Figure 4.3, we plot the PMs versus the observed values for the 20 sensitive counties. Some points are off the 45^o straight line with the parametric model lying just below the mixture model; the points from the individual area model are much below the 45^o line and points with smaller observed values are closer to

the 45^o. This again indicates that the individual area model gives poor AREs. The individual area model performs badly on the outliers.

In Figure 4.4, we compare the three models using the PSDs of the 20 sensitive counties. The best fitted straight line through the points for the individual area model is shown [intercept = 7.27, slope = 0.222, $R^2 = 0.971$]. Again the individual area model has the largest PSDs and the mixture model generally has the lowest PSDs; the outliers have much smaller PSDs under the mixture model than the parametric model, showing the strength of the mixture model.

Figure 4.2 Plots of the county posterior coefficient of variations (PCVs) versus the observed values by model

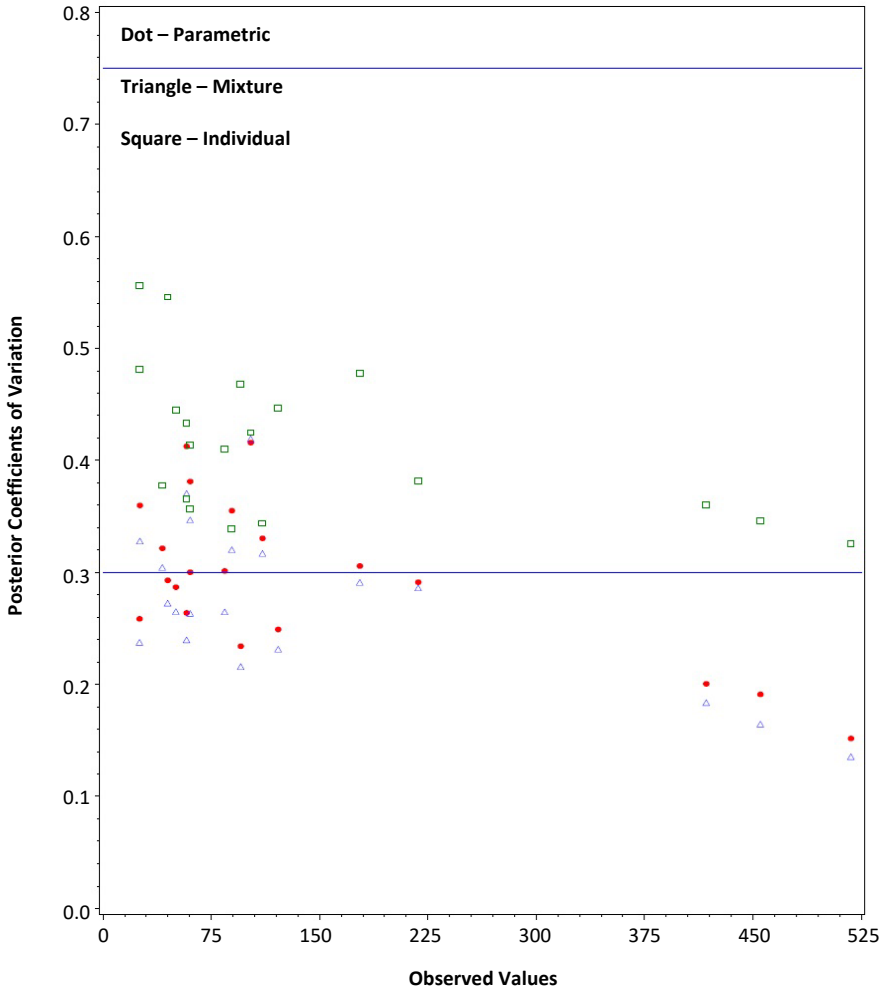
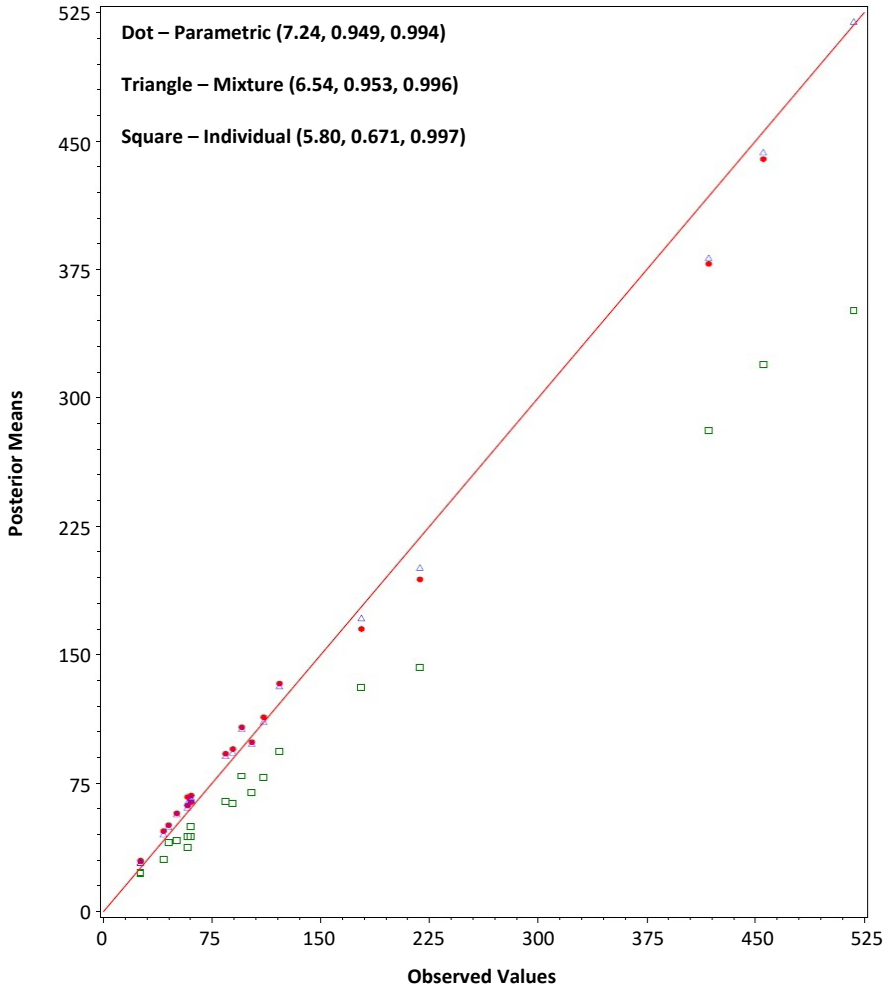
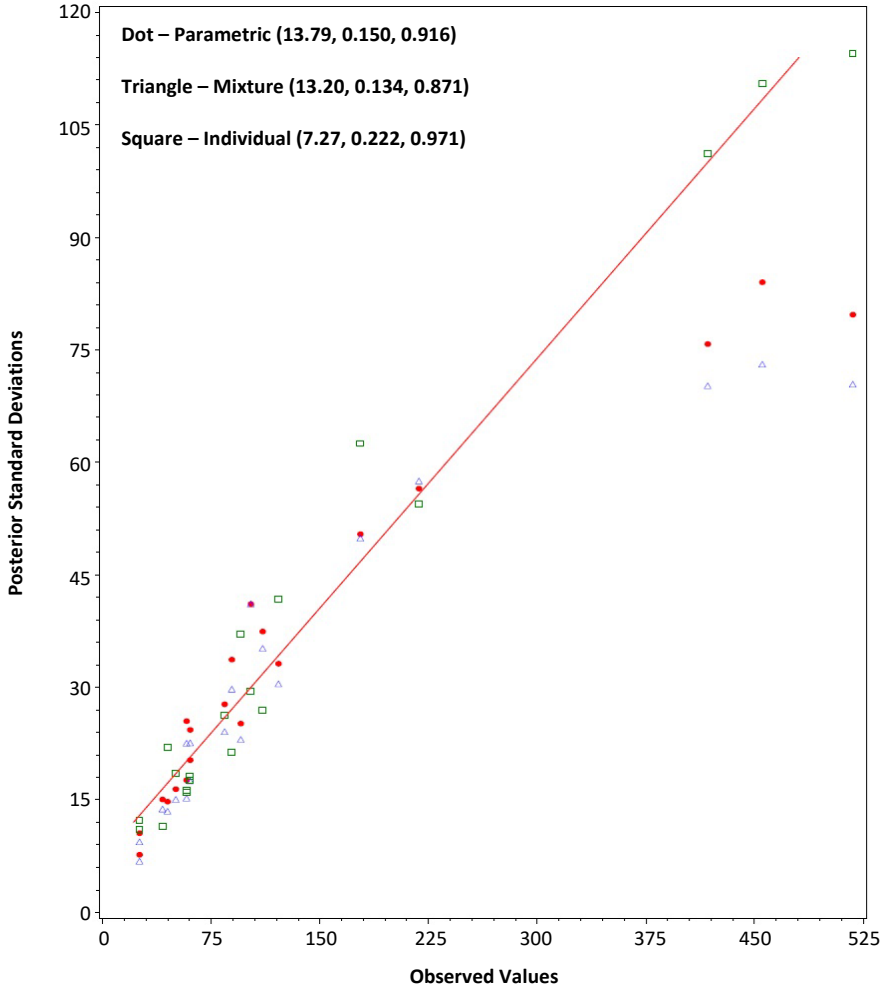


Figure 4.3 Plots of the county posterior means (PMs) versus the observed values by model



Based on these results (both tabular and graphical), it is therefore necessary to fit the small area models, especially the robust mixture model. It is not sensible to model each county separately and it is necessary to incorporate the similarity in the farms within a county and the counties within a state. The mixture model gives the best trade-off between security and utility; one important reason for this is that it is able to accommodate counties with very different acres (some small and some large, outliers) by automatically clustering the counties. In addition, many agencies are concerned with precision and the mixture model does well here also; both small area models beat the individual area model.

Figure 4.4 Plots of the county posterior standard deviations (PSDs) versus the observed values by model



Finally, we present Tables 4.4, 4.5 and 4.6, where we further compare the three models with respect to security and utility (both global and local). Comments are given in the note to each table, but we make general comments about each table here. In Table 4.4, we compare security based on the total budget, ϵ. We can see that the individual area model shows that a typical county has the highest security, but it is not reliable. On this same measure the parametric model and the mixture model are similar. In Table 4.5, we compare the AREs (local utility), and it is clear that the mixture model is the best, and the individual area model is the worst. However, all models show strong global utility based on the p-values and the pMSEs. In Table 4.6, we compare the three models using the 1,000 Gibbs samples. Based on five-number summaries, it may be possible to release some of these synthetic datasets (not just posterior means) if the original data are real; see Rubin (1993) and Raghunathan, Reiter and Rubin (2003).

Table 4.4
Comparison of security of the three models (parametric - P, mixture - M and individual - I) for sensitive counties

Model	PM	PSD	NSE	PCV	HPDI
P	2.204	0.044	0.001	0.020	(2.112, 2.250)
M	2.206	0.043	0.001	0.019	(2.112, 2.250)
I-T	0.811	0.576	0.016	0.710	(0.034, 1.983)
I-C	1.863	0.284	0.011	0.152	(1.326, 2.250)

Note: Entrees are summaries for the differential privacy parameter, ϵ, the total budget. I-T refers to a typical county and I-C refers to combined counties. P and M are very similar with strong security. A typical county (I-T) appears secure but it is unreliable (PCV is 0.710). When all counties are combined (I-C) under parallel composition, the budget is comparable (slightly smaller with larger PSD and PCV) than the two small area models. Highest posterior density interval (HPDI); numerical standard error (NSE); posterior coefficient of variation (PCV); posterior mean (PM); posterior standard deviation (PSD).

Table 4.5
Comparison of local utilities (AREs) of the three models (parametric - P, mixture - M and individual - I) for sensitive counties

Model	H	M	L	P
P	4	7	9	0
M	6	8	6	0
I	0	1	8	11

Note: Entrees are the number of sensitive counties with high (H), medium (M), low (L) and poor (P) utilities. The mixture model is better than the parametric model, and the individual model is not as good as these two models. For the parametric, mixture and individual models, the global utilities (as measured by the pMSEs) are 0.007, 0.008, 0.008 and (as measured by the p-values of WSRT) are 0.951, 0.938, 0.951 respectively, and therefore, all models show strong global utility. Absolute relative error (ARE); predictive mean squared errors (pMSEs); Wilcoxon signed rank test (WSRT).

Table 4.6
Comparison of global utilities of the three models (parametric - P, mixture - M and individual - I) for sensitive counties over the 1,000 Gibbs samples

Model	Min	Q_1	Q_2	Q_3	Max
a. P-values of WSRT					
P	0.867	0.927	0.942	0.957	0.996
M	0.856	0.928	0.944	0.956	0.999
I	0.892	0.931	0.941	0.952	0.998
b. pMSE, Snoke, Raab, Nowok, Dibben and Slavkovic (2018)					
P	0.000	0.000	0.001	0.003	0.030
M	0.000	0.000	0.001	0.003	0.020
I	0.006	0.007	0.009	0.013	0.058

Note: Entrees are five number summaries over the 1,000 Gibbs samples. All models have very strong global utility. If the data are real, not simulated, it may be possible to release some of these Gibbs samples because the smallest p-values of the WSRT are large and the largest pMSEs are not too large. Predictive mean squared errors (pMSEs); Wilcoxon signed rank test (WSRT).

5. Concluding remarks

We have constructed a hybrid Bayesian method, which includes a differentially private mechanism, to mask Census county totals for a U.S. state on acreage of a commodity using simulated farm level data. We have included a scale factor to the Laplace distribution, and a feature so that the overall budget in the modeling is ϵ as would normally be specified in the definition of differential privacy. Finally, we have applied two small area models, the parametric model and the novel robust mixture model to COM of STAT

in TAB; these are compared with an individual area model. It is important to separate out the sensitive counties. We have not looked at (ϵ, δ) - differential privacy, but it is possible that a similar methodology can be developed. Our intention in this paper is to show how we might proceed with more complex data structures, but our approach is important in many practical applications. We have provided a general Bayesian methodology that can be used for several problems with more complex models (e.g., linked tables and hierarchical tables).

Another novelty in this paper is the specification of the prior for ϵ. It is worth noting that we do not need to put a prior on ϵ, but this does not fit well into the Bayesian paradigm. We can certainly specify ϵ at values such as $\epsilon =$ 1, 2, as is common practice, but then we will need to perform a sensitivity analysis, and this is an additional burden. However, one can put a discrete prior on ϵ, $P(\epsilon = 1) = P(\epsilon = 2) = \frac{1}{2}$, in the same manner as we have done in this paper. This appears to make some sense within the Bayesian paradigm. Of course, the outputs are secured under differential privacy, but utility may not be all that good especially for very small counties.

We mention two possible extensions, which can potentially increase utility. First, it is possible to extend this approach to $r \times c$ tables with r counties and $c \geq 2$ commodities. In this work, we have done the case with $c = 1$ (i.e., a single commodity). Now, we want the margins of the masked county totals match the observed marginal totals, which can be sensitive though. However, the sparseness of the data is expected to cause difficulties; this sparseness can be avoided if only the major commodities are studied. Second, to improve utility we can add a spatial structure over the counties. We can use a conditional auto-regressive (CAR) model on the μ_i in our model, but this needs an incidence matrix for the counties of STAT with a particular commodity, COM. Both extensions are practically useful, but they are not straightforward. The Bayesian method for small areas in this paper serves as a template for more complex data structures.

At the request of a reviewer, to be concrete, we show how to generalize our method to data with covariates and where the study variable is binary (e.g., the U.S. Census Bureau and other agencies are also interested in percentages). We give a brief description of our parametric model with covariates such as sex of the operator, farm size, land tenure, crop diversity and irrigation, where the study variable is still harvested acres. Further generalization to percentages (binary data) is described in Appendix C. We assume the covariates are $\mathbf{x}_{ij}, j = 1, \ldots, n_i, i = 1., \ldots, \ell,$ and the study variable, y_{ij} for acres, is weighted as discussed for the three models. Both $(\mathbf{x}_{ij}, y_{ij})$ are microdata and these are not released. In the same way, we obtain the suppressed counties and the non-suppressed counties (non-sensitive and sensitive). Then, for the population model, we can assume that

$$y_{ij} \overset{\text{ind}}{\sim} \text{Normal}\,(\mathbf{x}'_{ij}\boldsymbol{\beta}, \sigma^2), j = 1, \ldots, n_i, i = 1, \ldots, \ell,$$

and for the sample model,

$$y_{ij} \overset{\text{ind}}{\sim} \text{Normal}\,(\mathbf{x}'_{ij}\boldsymbol{\beta} + (2r_i - 1)\, t_i, \sigma^2), j = 1, \ldots, n_i,$$

$$r_i \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{1}{2}\right), t_i \overset{\text{ind}}{\sim} \text{Gamma}\,(1, a_{oi}\epsilon), i = 1, \ldots, \ell,$$

$$\epsilon \sim \text{Gamma}\,(1,1), a_o < \epsilon < b_o, \pi(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2},$$

where the a_o, b_o and the a_{oi} are specified as described in the paper. The method to compromise a degree of security for an increase in utility goes true almost exactly as discussed in the paper: Run the Gibbs sampler, do the prediction and the benchmarking.

Finally, as indicated by the editor, we list seven caveats of our paper. These are mainly concerned with attempts to provide a trade-off between security and utility.

1. Parallel composition of differential privacy is not available in small area estimation. So to move forward, we have used a simple weighting scheme to partition the total budget among the counties.
2. We need to partition the non-suppressed counties into non-sensitive and sensitive counties, and sensitive counties need specific protection. It is good that all non-suppressed counties are fit to the models, but only the sensitive counties are benchmarked. It may be possible to make an adjustment to the models for the non-sensitive counties by using an offset.
3. We need to bound the AREs to get acceptable utility and precision. Using an upper bound less than unity (bigger than 0.25 is poor utility) appears to be reasonable. As we stated in the paper, the compromise between utility and security cannot be avoided.
4. The target formed by the total of the observations when the sensitive counties are benchmarked can be sensitive. In this case, we can use an alternative masking procedure to find a masked target, in which the ARE is say 5%.
5. While we have worked with county totals, which are the quantities of interest, it is possible to do the analysis using county averages (farm level data), which are less sensitive. We can then predict the county averages, and therefore the county totals, and with good security and utility, these can be released.
6. We can improve the prior specification of ϵ by borrowing the interval, $(0, b_o)$, constructed by Lee and Clifton (2011) using the observed confidential data. Then, a truncated exponential prior (or a uniform prior) can be used for ϵ. The difficulty here is that the probability that a participant is identified in the data set is required, and this is not practical.
7. Benchmarking using raking (proportional allocation) can be improved using a model-based analysis, but because of space, we do not present it in this paper.
8. We can obtain a good value of global sensitivity using a model-based analysis, but it is not necessary to present the details in this paper.

Disclaimer and Acknowledgements

All analyses presented in this paper were conducted using simulated data designed to mimic the structure and characteristics of the Census of Agriculture for illustrative purposes. No confidential or actual Census of Agriculture microdata were used in this study. The simulated dataset represents a hypothetical state with 75 counties and was generated solely for methodological demonstration. The findings and conclusions in this preliminary publication have not been formally disseminated by the U.S. Department of Agriculture and should not be construed to represent any agency determination or policy. This research was supported in part by the intramural research program of the U.S. Department of Agriculture (USDA), National Agriculture Statistics Service (NASS). The work of Balgobin Nandram was completed as a Senior Research Scientist on an IPA (Intergovernmental Personnel Act) at NASS. The authors thank many internal reviewers at NASS (Yang Cheng, Denise Abreu, Valbona Bejleri and Luca Sartore) for their comments, which led to a much improved manuscript. We also thank the two referees for their informative comments, and the Editor for encouragement.

Appendix

A. Laplace differential privacy mechanism

Let ϵ denote a positive real number, representing the privacy budget, and let A denote a randomized mechanism (e.g., Laplace distribution), mapping the data set, D, to $A(D)$. Let D_1 and D_2 differ on a single element (one farm or one county). Dwork (2006) called A ϵ - differential, if for all subsets S of $A(D)$,

$$P(A(D_1) \in S) \leq e^{\epsilon} P(A(D_2) \in S), \epsilon \geq 0,$$

where $P(\cdot)$ is the randomization distribution; also see Dwork and Roth (2014). It is also true that by symmetry,

$$P(A(D_2) \in S) \leq e^{\epsilon} P(A(D_1) \in S).$$

Therefore, letting $R = \frac{P(A(D_1) \in S)}{P(A(D_2) \in S)}$, we have $-\epsilon \leq \log(R) \leq \epsilon$.

The differential privacy mechanism has the property of composition which includes parallel, sequential and group compositions. It is robust to posterior processing (McSherry, 2009) if the observed confidential data are not used in post-processing; expert opinion can be used; see Williams and Bowen (2023) for a review.

The Laplace mechanism adds noise from a Laplace distribution. Let Y denote the noise, where $g(y) = \lambda e^{-\lambda|y|}, -\infty < y < \infty$; we note that $\text{Var}(Y) = \frac{2}{\lambda^2}$. We add y to $f(D)$ to get

$$z = f(D) + y,$$

where $f(D)$ is a possible query (may be the original data) and the noise, y, is $z - f(D)$. Then, for any z,

$$R = \frac{g\{z - f(D_1)\}}{g\{z - f(D_2)\}},$$

and under the Laplace mechanism,

$$R = \exp\left\{\lambda\left\{\left|z - f(D_2)\right| - \left|z - f(D_1)\right|\right\}\right\}.$$

Now, using the reversed triangular inequality,

$$R \leq \exp\left\{\lambda\left\{\left|f(D_2) - f(D_1)\right|\right\}\right\} \leq \exp\{\lambda\Delta(f)\},$$

where $\Delta(f) = \text{Sup}_{D_1,D_2}\left|f(D_2) - f(D_1)\right|$ is the sensitivity of the function $f(\cdot)$, and $\lambda\Delta(f)$ can be interpreted as the differential privacy parameter, ϵ. Therefore, it is worth noting that for the Laplace mechanism, we must have $\lambda = \frac{\epsilon}{\Delta(f)}$ for A to be an ϵ - differential private algorithm. Because both ϵ and λ are unknown, it is a difficult problem to choose the appropriate value of ϵ (e.g., see Lee and Clifton, 2011), but choosing a small value of ϵ is sensible.

B. The P-percent sensitivity rule

Consider a single county of n farms with acreages, $y_1,\ldots,y_n$, which we now order from smallest to largest, $y_{(1)} \leq \ldots \leq y_{(n)}$ and let $T = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_{(i)}$. The p-percent rule says that a county is sensitive if

$$T - (y_{(n-1)} + y_{(n)}) \leq \frac{p}{100} y_{(n)},$$

where $0 < p < 100$. Equivalently, a county is non-sensitive if $T - (y_{(n-1)} + y_{(n)}) > \frac{p}{100} y_{(n)}$. This rule simply states that a county is sensitive if $\sum_{i=1}^{n-2} y_{(i)} \leq \frac{p}{100} y_{(n)}$ or non-sensitive if $\sum_{i=1}^{n-2} y_{(i)} > \frac{p}{100} y_{(n)}$ and $n \geq 3$.

This partition of the farms into two sets shows that if the farm with $y_{(n-1)}$ can estimate $y_{(n)}$ very well and the sum of the remaining counties (not the largest and second largest) is small, a county will be sensitive. An attacker with his value $y_{(n-1)}$ and T can estimate $y_{(n)}$ in the interval,

$$\frac{T - y_{(n-1)}}{1 + \frac{p}{100}} \leq y_{(n)} \leq T - y_{(n-1)}.$$

C. Percentages and logistic regression

As we are interested in percentages, it is natural to consider logistic regression, and we denote the microdata by $(w_{ij}, \mathbf{x}_{ij}, y_{ij}), j = 1,\ldots,n_i, i = 1,\ldots,\ell$, where y_{ij} are the binary study variables. We will apply the Census weights to the covariates, $\tilde{\mathbf{x}}_{ij} = w_{ij}\mathbf{x}_{ij}$, instead of the binary variables. This is technique we use for weighted logistic regression (e.g., Yang, Nandram and Choi, 2023).

The $3+$ rule is used the same way to suppress counties with fewer then 3 farms. We can now partition the non-suppressed counties into non-sensitive and sensitive counties using the propensity scores from the model,

$$y_{ij} \mid \boldsymbol{\beta} \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{e^{\tilde{\mathbf{x}}_{ij}'\boldsymbol{\beta}}}{1+e^{\tilde{\mathbf{x}}_{ij}'\boldsymbol{\beta}}}\right), j=1,\ldots,n_i, i=1,\ldots,\ell.$$

We can compute maximum likelihood estimator (same as posterior mode under a flat prior for $\boldsymbol{\beta}$), $\hat{\boldsymbol{\beta}}$. Then, the propensity scores are

$$\pi_{ij} = \frac{e^{\tilde{\mathbf{x}}_{ij}'\hat{\boldsymbol{\beta}}}}{1+e^{\tilde{\mathbf{x}}_{ij}'\hat{\boldsymbol{\beta}}}}, j=1,\ldots,n_i, i=1,\ldots,\ell.$$

For the i^{th} county, we apply the p-percent rule to the π_{ij} to decide whether it is non-sensitive or sensitive.

For the population model, we assume

$$y_{ij} \mid \boldsymbol{\beta} \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{e^{\tilde{\mathbf{x}}_{ij}'\boldsymbol{\beta}}}{1+e^{\tilde{\mathbf{x}}_{ij}'\boldsymbol{\beta}}}\right), j=1,\ldots,n_i, i=1,\ldots,\ell.$$

For the sampling model, we assume

$$y_{ij} \mid \boldsymbol{\beta} \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{e^{\tilde{\mathbf{x}}_{ij}'\boldsymbol{\beta}+(2r_i-1)t_i}}{1+e^{\tilde{\mathbf{x}}_{ij}'\boldsymbol{\beta}+(2r_i-1)t_i}}\right), j=1,\ldots,n_i, i=1,\ldots,\ell,$$

$$r_i \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{1}{2}\right), t_i \overset{\text{ind}}{\sim} \text{Gamma}(1, a_{oi}\epsilon), i=1,\ldots,\ell,$$

where a_{oi} are the scale parameters. These a_{oi} can be chosen to be

$$a_{oi} = \frac{n_i / t_{oi}}{\sum_{i'=1}^{\ell} n_{i'} / t_{oi'}}, \quad \text{where} \quad t_{oi} = \sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n_i}} \quad \text{and} \quad \hat{p}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}, i=1,\ldots,\ell.$$

Finally, as in the models stated in this paper, with independence of $\boldsymbol{\beta}$ and ϵ, we assume

$$\pi(\beta) = 1, \ \epsilon \sim \text{Gamma}(1,1), a_o < \epsilon < b_o,$$

where a_o and b_o are specified. This model can be fit using a systematic-scan Gibbs sampler via the Pólya-Gamma data augmentation with some blocking; see Polson, Scott and Windle (2013).

For prediction, we sample

$$z_{ij} \sim \text{Bernoulli}\left(\frac{e^{\tilde{\mathbf{x}}_{ij}'\boldsymbol{\beta}}}{1+e^{\tilde{\mathbf{x}}_{ij}'\boldsymbol{\beta}}}\right), j=1,\ldots,n_i, i=1,\ldots,\ell,$$

to provide acceptable utility, we use the restriction (as in the paper),

$$(1-a)\,\bar{y}_i < \bar{z}_i < (1+a)\,\bar{y}_i, \ 0<a<1, \ \bar{z}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} z_{ij}, i=1,\ldots,\ell,$$

where a is specified. The $\overline{z}_i$ are the proportions we need to benchmark and the $\overline{y}_i$ are the observed proportions. Then, the benchmarked county proportions (multiply by 100 for percentages) are

$$P_i = \overline{z}_i \frac{\sum_{i=1}^{\ell} n_i \overline{y}_i}{\sum_{i=1}^{\ell} n_i \overline{z}_i}.$$

Of course, we are actually doing benchmarking of the sensitive counties to their observed total (this total may be sensitive; see text).

References

Awan, J. and Slavkovic, A. (2019). *Differentially Private Inference for Binomial Data*. Available at: https://arXiv:1904.00459v1[math.ST], 31 March 2019, 1-39.

Bowen, C.M. and Liu, F. (2020). Comparative study of differentially private data synthesis methods. *Statistical Science*, 35(2), 280-307.

Bring, J. and Wang, Q. (2014). Comparison of different sensitivity rules for tabular data and presenting a new rule – The interval rule, (Ed., J. Domingo-Ferrer): PSD 2014, LNCS 8744, 36-47, Springer International Publishing Switzerland.

Cox, L.H. (1980). Suppression methodology and statistical disclosure control. *Journal of the American Statistical Association*, 75(370), 377-385.

Cox, L.H. (1995). Network models for complementary cell suppression. *Journal of the American Statistical Association*, 90, 1453-1462.

Cox, L.H., Kelly, J.P. and Patil, R.J. (2004). Computational aspects of controlled tabular adjustment: Algorithm and analysis. *The Next Wave in Computing, Optimization, and Decision Technologies*, 45-59.

Drechsler, J. (2023). Differential privacy for government agencies – Are we there yet? *Journal of the American Statistical Association*, 118(541), 761-773. https://doi.org/10.1080/01621459.2022.2161385.

Duncan, G.T. and Lambert, D. (1986). Disclosure-limited data dissemination. *Journal of the American Statistical Association*, 81(393), 10-18.

Duncan, G.T. and Lambert, D. (1989). The risk of disclosure of microdata. *Journal of Business and Economic Statistics*, 7(2), 207-217.

Dwork, C. (2006). Differential privacy. *Microsoft Research*, 1-12.

Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference (TCC),* (Eds., S. Halevi and T. Rabin), Springer, 265-284.

Dwork, C., Naor, M., Reingold, O., Rothblum, G. and Vadhan, S. (2009). On the complexity of differentially private data release: Efficient algorithms and hardness results. *Proceedings of the forty-first annual ACM Symposium on Theory of Computing*, 381-390. https://doi.org/10.1145/1536414.1536467.

Dwork, C. and Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4), 211-407. https://doi.org/10.1561/0400000042.

Evans, T., Zayata, L. and Slanta, J. (1996). Using noise for disclosure limitation of establishment tabular data. Proceedings: *Annual Research Conference and Technology Interchange*, Bureau of the Census, 65-68.

Geweke, J. (1992). Evaluating the accuracy of the sample-based approach to calculation of posterior moments. In *Bayesian Statistics*, (Eds., J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), 4, 169-193.

Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., Narayan, A., Pierce, B.C. and Roth, A. (2014). Differential privacy: An economic method for choosing epsilon. Available at: https://arXiv:1402.3329V1 and https://do1.org/ 10.48550/arXiv.1402.332, February 13, 2014, 1-29.

Hu, J. and Bowen, C.M. (2024). Advancing microdata privacy protection: A review of synthetic data methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 16(1), e1636, 1-19.

Hu, J., Williams, M.R. and Savitsky, T.D. (2025). Mechanisms for global differential privacy under Bayesian data synthesis. *Statistica Sinica*, 35(1), 563-584.

Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453), 161-172.

Janicki, R., Holan, S.H., Irimata, K.M., Livsey, J. and Raim, A. (2024). *Bayesian Methods to Improve the Accuracy of Differentially Private Measurements of Constrained Parameters*. Available at: https://arXiv:2406.04448v1[Stat.ME], 6 June 2025, 1-7.

Lee, J. and Clifton, C. (2011). How much is enough? Choosing ϵ for differential privacy. *Proceedings of the International Conference on Information Security*, Springer, Berlin, Heidelberg, 325-340.

Little, R.J.A. (1993). Statistical analysis of masked data. *Journal of Official Statistics*, 9(2), 407-426.

Manandhar, B. and Nandram, B. (2021). Hierarchical Bayesian models for continuous and positively skewed data from small areas. *Communications in Statistics-Theory and Methods*, 50(4), 944-962.

Manandhar, B. and Nandram, B. (2025). Hierarchical Bayesian models for small area estimation with GB2 distribution. *Journal of Applied Statistics*, 52(1), 1-30.

McSherry, F.D. (2009). Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*. https://doi.org/10.1145/1559845.1559850.

Nandram, B., Cruze, N.B. and Erciulescu, A.L. (2023). Bayesian small area models under inequality constraints with benchmarking and double shrinkage. *Survey Methodology*, 49(2), 517-546. Available at: https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2023002/article/00004-eng.pdf.

Nandram, B., Toto, M.C.S. and Choi, J.W. (2011). A Bayesian benchmarking of the Scott-Smith model for small areas. *Journal of Statistical Computation and Simulation*, 81(11), 1593-1608.

Nandram, B. and Yu, Y. (2019). Bayesian analysis of a sensitive proportion for a small area. *International Statistical Review*, 87, S1, S104-S120. https://doi.org/10.1111/insr.12286.

Polson, N.G., Scott, J.G. and Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association,* 108(504), 1339-1349.

Poole, W.K. (1974). Estimation of the distribution function of a continuous type random variable through randomized response. *Journal of the American Statistical Association*, 69(348), 1002-1005.

Raghunathan, T.E., Reiter, J.P. and Rubin, D.B. (2003). Multiple imputation for statistical disclosure limitation. *Journal of Official Statistics*, 19(1), 1-16.

Reiter, J.P. (2019). Differential privacy and federal data releases. *The Annual Review of Statistics and its Application*, 6, 85-101.

Rinott, Y., O'Keefe, C.M., Shlomo, N. and Skinner, C. (2018). Confidentiality and differential privacy in the dissemination of frequency tables. *Statistical Science*, 33(3), 358-385.

Rubin, D.B. (1993). Discussion: Statistical disclosure limitation. *Journal of Official Statistics*, 9(2), 461-468.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639-650.

Snoke, J., Raab, G.M., Nowok, B., Dibben, C. and Slavkovic, A. (2018). General and specific utility measures for synthetic data. *Journal of the Royal Statistical Society*, Series A, 181(3), 663-688.

Tan, A. and Hobert, J.P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *Journal of Computational and Graphical Statistics*, 18(4), 861-878.

Williams, A.R. and Bowen, C.M. (2023). The promise and limitations of formal privacy. *Computational Statistics*, 15(2), e1615, 1-17.

Yang, L., Nandram, B. and Choi, J.W. (2023). Bayesian predictive inference under nine methods for incorporating survey weights. *International Journal of Statistics and Probability*, 12(1), 33-53.

Zhang, Z., Rubinstein, B.I.P. and Dimitrakakis, C. (2016). On the differential privacy of Bayesian inference. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), Association for the Advancement of Artificial Intelligence* (Available at: https://www.aaai.org), 2365-2371.