RESEARCH REPORT SERIES
(Statistics #2022-05)

Evaluation of Bayesian Hierarchical Models of Differentially Private Data Based on an Approximate Data Model

Kyle M. Irimata[1],
Andrew M. Raim[1],
Ryan Janicki[1],
James A. Livsey[1],
Scott H. Holan[2,3]

[1]Center for Statistical Research and Methodology, U.S. Census Bureau
[2]Department of Statistics, University of Missouri
[3]Office of the Associate Director for Research and Methodology, U.S. Census Bureau

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: September 30, 2022

Evaluation of Bayesian Hierarchical Models of Differentially Private Data Based on an Approximate Data Model

Kyle M. Irimata[a], Andrew M. Raim[a], Ryan Janicki[a], James A. Livsey[a], and Scott H. Holan[b,c]

[a]Center for Statistical Research and Methodology, U.S. Census Bureau
[b]Department of Statistics, University of Missouri
[c]Office of the Associate Director for Research and Methodology, U.S. Census Bureau

2022-09-30

Abstract

Official statistical agencies have an obligation to protect the confidentiality of their respondents. Traditionally, the task of disclosure avoidance has consisted of a variety of different approaches, including cell suppression and data swapping, among others. More recently, official statistical agencies (e.g., the U.S. Census Bureau) have employed differential privacy (DP) methods to facilitate DA. A major component of this method of DA involves the addition of noise to the observed measurements from an appropriate DP noise distribution based on a pre-specified privacy-loss budget. Consequently, it may be desirable to develop statistical models for the DP noisy measurements which can be used to produce more accurate synthetic data releases. One natural way to achieve this is using a Bayesian hierarchical model with an approximation to the DP noise distribution at the data level of the model hierarchy. Through model-based and empirical simulations, this paper evaluates the accuracy of this modeling approach under various settings of the DP noise distribution, the privacy-loss budget, and different approximations. We show that the Gaussian distribution provides a suitable, and computationally tractable approximation to Laplace and Discrete Gaussian noise mechanisms when the magnitude of the noise mechanisms is not too large.

1 Introduction

The U.S. Census Bureau and other official statistical agencies employ statistical disclosure avoidance methodology as a means of protecting the confidentiality of their respondents. To date, many different disclosure avoidance methods have been developed, including data swapping, cell suppression, and synthetic data methods, among others. Although the U.S. Census Bureau utilized various disclosure avoidance techniques in previous decennial censuses, it has been reported that these techniques were no longer suitable for protecting the 2020 decennial census. The 2010 U.S. decennial census counted a total population of over 308 million persons, and published at least 7.7 billion statistics from the collected data, or over 25 statistics per person. Experiments conducted by the Census Bureau using the 2010 decennial census data showed that the confidential micro-data could be accurately reconstructed from a subset of the published census statistics (Abowd, 2018), in potential violation of Title 13 U.S.C., which provides confidentiality protections for individual respondents[1].

Due to these concerns, the U.S. Census Bureau is implementing differential privacy (DP) for many of its 2020 decennial census data products (Abowd et al., 2021). In this case, DP provides mathematical privacy guarantees of individual responses by adding noise to summary statistics. The amount of additive noise is

Disclaimer: This article is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the authors and not those of the U.S. Census Bureau.

[1] www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html

parameterized by a privacy budget, typically denoted by $\epsilon > 0$. Small values of ϵ near zero greatly reduce the risk that an individual respondent can be identified, at the expense of published statistics that are less accurate with respect to the data as collected. Conversely, larger values of ϵ increase the risk of unintended disclosure, but ensure accurate statistics.

Although DP provides mathematical privacy guarantees, there is also a need to balance the privacy-loss budget to simultaneously provide ample privacy protection while still providing data that are fit for use (depending on the use-case). "Optimizing" this privacy-loss budget results in tables in which aggregate cells need to be tabulated in order to avoid unnecessarily allocating privacy-loss budget to cells that could alternatively be calculated (e.g., directly through aggregation). This optimization must also balance the expenditure of the privacy-loss budget on queries of interest and overall query sensitivity, while also ensuring the privacy-loss budget is not split into too many pieces.

The conflict between the need for increased privacy protections for census respondents and the desire to maintain the gold-standard quality of census data products has led to increased research on using model-based methods to increase the precision of differentially private measurements. Such model-based methods may incorporate auxiliary information that does not expend any additional privacy loss budget to increase the precision of the tabulations that are being disseminated, and to provide tabulations that may not otherwise be released. One appealing modeling option is a Bayesian hierarchical model (BHM), where the data model conforms to the known DP distribution being used by the agency, while a latent Gaussian process model (LGP) is utilized to model the true underlying latent counts. Publicly available auxiliary information, such as past decennial census data and aggregated sample survey data, such as American Community Survey (ACS) data, can be easily incorporated in a BHM as regression predictors. Also, various sources of dependence, such as spatial dependencies and/or multivariate dependencies within decennial census tabulations, can be leveraged to improve precision of published estimates.

The noise mechanism used in the application of differential privacy is assumed to be known, and can naturally be incorporated into a BHM, along with a mixed effects model for the true counts which are modeled on the logarithmic scale. However, there are unique challenges to model development for differentially private measurements. The main difficulty is that one of the most popular distributions used for generating differentially private measurements, the Laplace distribution, presents computational challenges when directly used in the data model of a BHM. We also consider modeling data which has been protected by a mechanism based on the discrete Gaussian distribution. We address both situations by proposing a continuous Gaussian distribution to approximate the noise mechanism. Post-processing may be used in practice to satisfy data quality constraints, such as ensuring non-negative population counts and hierarchical consistency in a group of related tabulations (e.g., Abowd et al., 2022). Accounting for such adjustments in a model may be significantly more complicated than the DP mechanism itself. Therefore, we assume throughout this work that modeling is carried out after DP but before any post-processing, as a means of potentially improving the accuracy of published tabulations.

In this paper, we demonstrate the benefits of using a BHM to improve the accuracy of tabulations treated with DP, as compared to utilizing the noisy measurements directly. Several variations of lognormal models are demonstrated as practical examples of LGPs. We critically evaluate the performance of a Gaussian distribution in place of the corresponding DP data distribution in a BHM with respect to computational time and accuracy. If the approximation is demonstrated to provide accurate tabulations, along with measures of uncertainty, then this would ultimately open the door for model-based tabulations to be produced in a massive production setting. Though this investigation is motivated by the Census Bureau's adoption of DP, the settings used and analyses conducted are not necessarily indicative of the approaches that will be used in production.

This paper proceeds as follows. Section 2 provides background on differential privacy and Bayesian hierarchical modeling. In particular, we discuss the popular data model, process model, and parameter model paradigm (Cressie and Wikle, 2015) and how the DP distribution is incorporated into the data model specification. Simulations using the Laplace mechanism for DP are provided in Section 3. In contrast, Section 4 provides simulation results in the context for which the discrete Gaussian distribution is the DP noise. Sections 5 and 6 provide the results of an empirical simulation study using Laplace noise and discrete

Gaussian noise respectively. Finally, Section 7 concludes.

2 The DP Data Model and Bayesian Hierarchical Models

2.1 A Motivating Dataset

The 2020 U.S. Decennial Census will be the first to use DP as the primary method of privacy protection.[2] At the time of this writing, the Public Law 94–171 File, which is used by states to conduct legislative redistricting, has been released to the public. Here, DP is implemented via the TopDown Algorithm (Abowd et al., 2022), although state-level population counts are held invariant so that exact enumerations are available for redistricting.[3] Other products, such as a Demographic and Housing Characteristics File and a Detailed Demographic and Housing Characteristics File are being prepared and are not yet fully specified.

For privacy-protected datasets, the sensitive underlying data will not be available to produce results which could potentially become public (Abowd, 2018). To study the effect of DP, we will consider tabulations from the 2010 decennial census which have been publicly released after privacy protection using non-DP methods. Regarding these data as exact enumerations, we may apply standard DP methods and consider models to make inference on the underlying enumerations through the protected data, though we do not apply nor investigate the effects of post-processing. This setting motivates the remainder of the paper. The release from the 2010 census contains a number of tabulations across several data products; we focus specifically on a single tabulation PCT1 "American Indian and Alaska Native Alone with One Tribe Reported for Selected Tribes" and three particular fields which represent counts of American Indian and Alaska Native (AIAN) groups: Cherokee, Choctaw, and Sioux. These are among the largest AIAN groups by total population at the national level, and are represented in other datasets that we will describe shortly. We consider county-level tabulations for Oklahoma and its surrounding states: Colorado, Kansas, Missouri, Arkansas, Texas, and New Mexico. Oklahoma was selected as a spatially coherent illustrative example because of the presence of Cherokee, Choctaw and Sioux persons. Counts for the 723 counties in this region are displayed in Figure 1, transformed to the log-scale using $f(x) = \log(x+1)$ to improve visualization and accommodate zero counts. Each group contains a number of counties with small or zero counts; this is especially seen for Sioux. The counties with smaller counts often appear to be clustered together, suggesting that there may be a spatial effect. The empirical densities shown in Figure 1d emphasize that the counts for each group are highly right-skewed and will influence model selection.

Regarding the 2010 data (with additional privacy protection by DP) as outcomes to be modeled, obvious sources of predictors are previous censuses and releases of ACS estimates. We therefore consider the 2000 census and 2009 ACS 5-year estimates. From the 2000 census, we make use of analogous data to 2010, via the fields for Cherokee, Choctaw, and Sioux from tabulation PCT1 "American Indian and Alaska Native Alone with One Tribe Reported for Selected Tribes". From the 2009 ACS release, we make use of 5-year period estimates from table B02005 "American Indian and Alaska Native Alone for Selected Tribal Groupings", using point estimates for Cherokee, Choctaw, and Sioux population counts. Margins of error (MOEs) for ACS estimates are also available but are not used in this work, though future work may consider incorporating measurement error models to utilize this information. Figure 2 displays scatterplots comparing the 2010 census counts to both 2000 census counts and ACS point estimates for each of the three groups. We notice a strong linear relationship between the 2000 census and 2010 census counts for all three groups. The relationship between the ACS estimates and 2010 census counts appears strongest for Cherokee, showcasing some outliers for larger Choctaw counts, and with more noticable deviation for the Sioux race group. This suggests that both 2000 census counts and ACS point estimates will be useful predictors, but the former may provide the most utility in formulating model-based predictions for unobservable sensitive data. We will revisit these data in Sections 5 and 6.

Data from the 2010 census, 2000 census, and 2009 ACS may be obtained in a number of ways. In

[2]Description of 2020 data products: https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/release/about-2020-data-products.html.

[3]https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files.html.

particular, the website https://data.census.gov is useful for browsing available data. The Census Bureau data API https://api.census.gov provides a web service to make programmatic requests for data. Users of R (R Core Team, 2022) can request data from the API within R using the tidycensus package (Walker and Herman, 2022). Supplementary materials included with this paper demonstrate the tidycensus approach.

2.2 Hierarchical models

Consider modeling privacy-protected data released by an agency. Denote $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ as sensitive data which is internal to the agency. A protected version of the data $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ is released to the public, where $Z_i = Y_i + \xi_i$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ is generated from a privacy protection mechanism which will be discussed in Section 2.3. In this application, these mechanisms are realized by univariate distributions which are fully known to the analyst including any parameters. Covariates $\boldsymbol{x}_i \in \mathbb{R}^d$ will be assumed to be available without privacy protection, given that non-DP methods were applied to past data products. We will consider three Bayesian hierarchical models based on log-normally distributed process models with added DP noise: an independent and identically distributed model (IID), a regression model (REG) based on a given covariate and independent errors, and a regression model whose errors are spatially dependent via conditional autoregression (CAR). These will be expressed as hierarchical models with three levels (Cressie and Wikle, 2015) a data model, a latent process model, and a parameter model.

To formulate the CAR model, suppose we are modeling a geographical domain consisting of n areal units. Let $\boldsymbol{A} = (a_{ij})$ be the $n \times n$ adjacency matrix with

$$a_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and areas } i \text{ and } j \text{ are adjacent,} \\ 0 & \text{otherwise,} \end{cases}$$

for $i, j \in \{1, \ldots, n\}$. Let $\boldsymbol{D} = \text{Diag}(\boldsymbol{A}\boldsymbol{1}_n)$ be an $n \times n$ diagonal matrix of adjacency counts $a_{1+}, \ldots, a_{n+}$ for each areal unit, where $\boldsymbol{1}_n$ is a vector of n ones. A random variable $(T_1, \ldots, T_n) \sim \text{N}(\boldsymbol{\mu}, \sigma^2(\boldsymbol{D} - \rho\boldsymbol{A})^{-1})$ if and only if each coordinate has the conditional distribution

$$[T_i \mid \boldsymbol{T}_{-i} = \boldsymbol{t}_{-i}] \sim \text{N}\left(\mu_i + \frac{\rho}{a_{i+}} \sum_{\ell=1}^{n} a_{i\ell}(t_\ell - \mu_\ell), \frac{\sigma^2}{a_{i+}}\right) \tag{2.1}$$

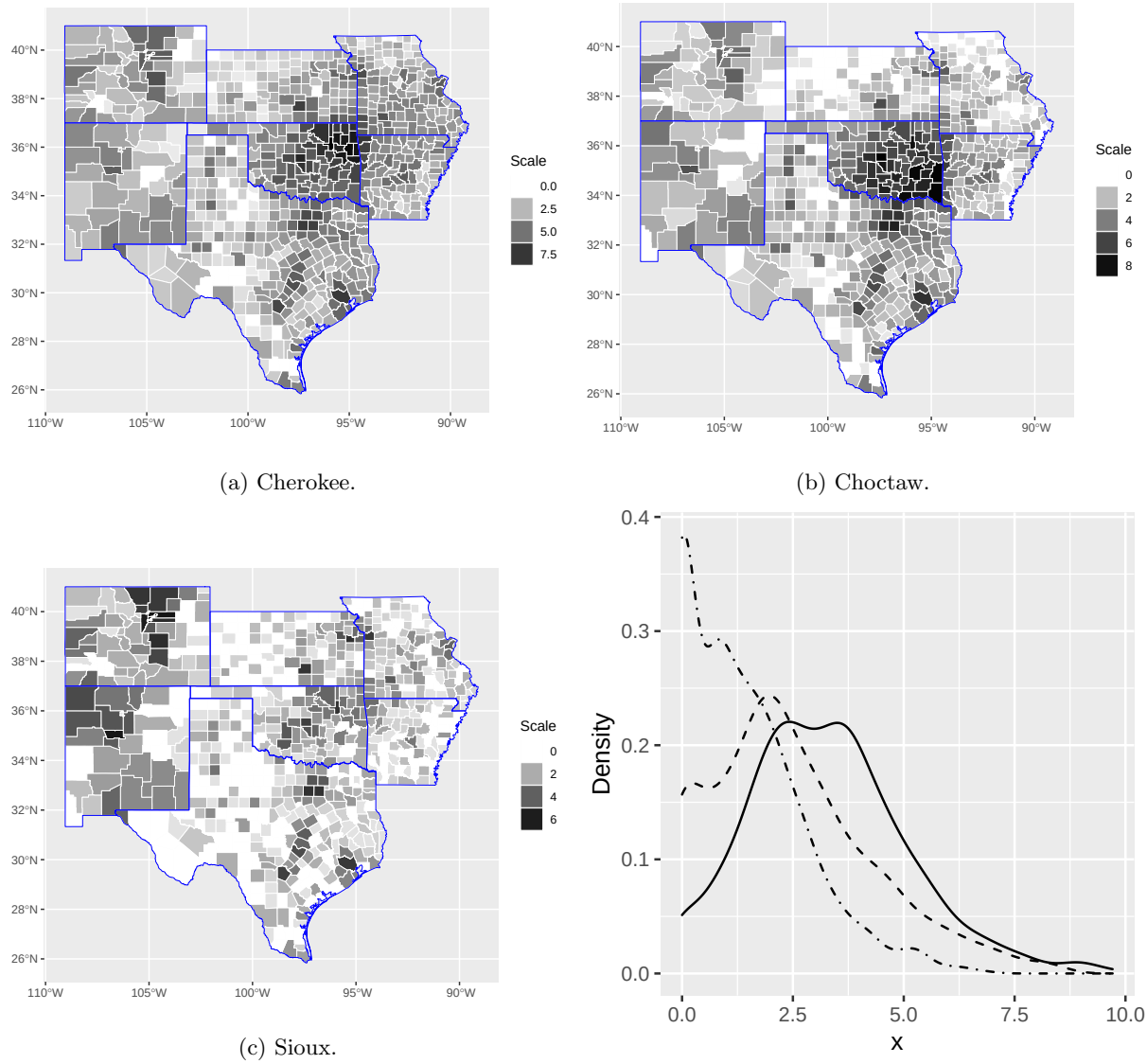
(e.g., see Cressie, 1993, Chapter 6), where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, σ^2, and ρ are parameters. The IID, REG, and CAR models assume the same data model

$$Z_i = Y_i + \xi_i, \quad i = 1, \ldots, n, \tag{2.2}$$

where the observed Z_i is a noisy observation based on latent Y_i with added agency noise ξ_i. The distribution of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ will be determined by the process model for each scenario:

$$\begin{aligned} \text{IID:} \quad & \log Y_i = \mu + \eta_i, \quad \eta_i | \sigma^2 \overset{\text{i.i.d.}}{\sim} \text{N}(0, \sigma^2), \\ \text{REG:} \quad & \log Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \eta_i, \quad \eta_i | \sigma^2 \overset{\text{i.i.d.}}{\sim} \text{N}(0, \sigma^2), \\ \text{CAR:} \quad & \log Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \eta_i, \quad \boldsymbol{\eta} | \sigma^2 \sim \text{N}(\boldsymbol{0}, \sigma^2(\boldsymbol{D} - \rho\boldsymbol{A})^{-1}), \end{aligned}$$

where $i = 1, \ldots, n$ and $\boldsymbol{\eta}^\top = (\eta_1, \ldots, \eta_n)$ represents the random error. For the CAR model, we assume a fixed value of $\rho < 1$ so that the matrix $\boldsymbol{D} - \rho\boldsymbol{A}$ is nonsingular. Alternatively, estimation of ρ can be avoided by assuming an intrinsic CAR (ICAR) dependence structure (Besag and Kooperberg, 1995). The vector $\boldsymbol{\theta}$ will refer to the unknown parameters in each setting; that is, $\boldsymbol{\theta} = (\mu, \sigma^2)$ in the IID model and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ in the REG and CAR models. The final level of the hierarchy—the parameter model—will vary across simulation settings, and will therefore be specified in those sections.



(a) Cherokee.

(b) Choctaw.

(c) Sioux.

(d) Cherokee (solid), Choctaw (dash), Sioux (dash-dot).

Figure 1: 2010 census counts—transformed to the log-scale using $f(x) = \log(x + 1)$ for visualization—of three AIAN race groups among 723 counties, and smoothed density.

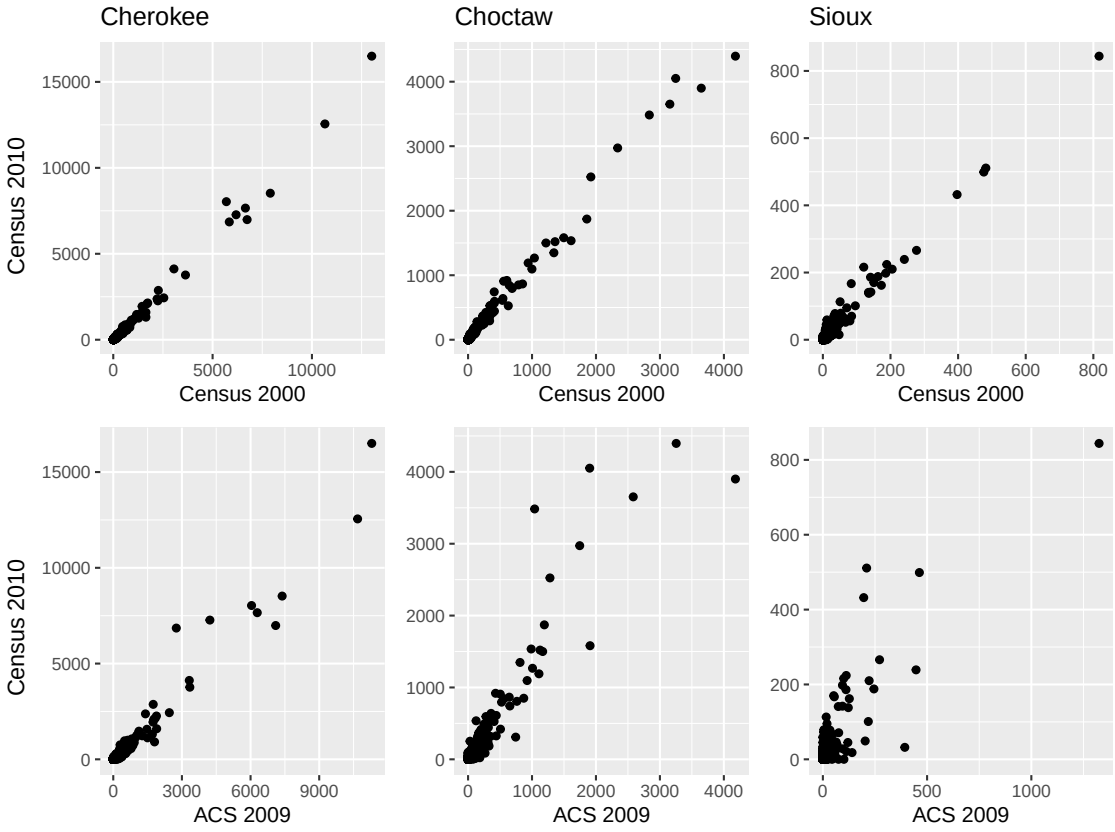


Figure 2: 2010 census counts versus auxiliary data for three AIAN race groups among 723 counties.

2.3 Differential Privacy Mechanisms

The IID, REG, and CAR models presented in Section 2.2 assume data model (2.2) where data are only released after applying additive agency noise $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$. In this section, we will briefly review several differential privacy mechanisms which can be used to generate $\boldsymbol{\xi}$. Each mechanism considers ξ_i to be independent and identically distributed, with a distribution depending on a known scale parameter λ_i, and absent any post-processing. An overview of differential privacy can be found in Dwork and Roth (2014).

For the remainder of this section, let the sensitive data procured by an agency be denoted by $\boldsymbol{x}$ and suppose $q(\boldsymbol{x}) \in \mathbb{R}^n$ is a vector of statistics to be released in a protected form. Here, $\boldsymbol{x}$ is assumed to be a histogram (or table) where each entry represents a number of records matching a unique combination of characteristics. The set of all possible histograms is denoted $\mathcal{X}$. The function q represents a query on a given histogram; e.g., given records $\boldsymbol{x}$ representing the population count in each state of the U.S. by age, one choice of $q(\boldsymbol{x})$ could produce the number of people aged 18 or younger in each state. We will consider several randomized noise mechanisms of the form

$$T(\boldsymbol{x}; q) = q(\boldsymbol{x}) + \boldsymbol{\xi}, \tag{2.3}$$

where the distribution of $\boldsymbol{\xi}$ is to be specified. To discuss differential privacy of such mechanisms, let $\|\boldsymbol{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ denote the l_p norm of $\boldsymbol{x} \in \mathcal{X}$ for $p \geq 1$. Let $\mathcal{D}_1 = \{(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{X} \times \mathcal{X} : \|\boldsymbol{x} - \boldsymbol{x}'\|_1 = 1\}$ represent pairs of histograms $\boldsymbol{x}$ and $\boldsymbol{x}'$ which differ by a single record. The l_p-sensitivity of a function $g : \mathcal{X} \mapsto \mathbb{R}^n$ is defined as

$$\Delta_p g = \max_{(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{D}_1} \|g(\boldsymbol{x}) - g(\boldsymbol{x}')\|_p.$$

We may now state one of the commonly used criteria for differential privacy.

Definition 2.1. Let $\epsilon > 0$ and let $q : \mathcal{X} \mapsto \mathbb{R}^n$ be a query. A randomized noise mechanism $T : \mathcal{X} \mapsto \mathcal{A}$ is said to satisfy ϵ-differential privacy if

$$\mathrm{P}(T(\boldsymbol{x}; q) \in A) \leq e^{\epsilon}\, \mathrm{P}(T(\boldsymbol{x}'; q) \in A) \tag{2.4}$$

for all measurable subsets $A \subseteq \mathcal{A}$ and all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ such that $\|\boldsymbol{x} - \boldsymbol{x}'\|_1 \leq 1$.

The privacy loss budget $\epsilon > 0$ is selected by the agency to quantify the level of protection to be received by the data. A small value of ϵ corresponds to a large amount of protection as well as a larger amount of noise in T. The following examples recall two noise mechanisms of the form (2.3). Suppose $q : \mathcal{X} \mapsto \mathbb{R}^n$ has finite sensitivity and $\boldsymbol{x} \in \mathcal{X}$.

Example 2.2. The Laplace distribution $\mathrm{Lap}(\lambda)$ is defined by the probability density function

$$f_{\mathrm{Lap}}(x; \lambda) = \frac{1}{2\lambda} e^{-|x|/\lambda}$$

The associated Laplace mechanism

$$T_{\mathrm{Lap}}(\boldsymbol{x}; q) = q(\boldsymbol{x}) + \boldsymbol{\xi}, \quad \xi_i \overset{\text{i.i.d.}}{\sim} \mathrm{Lap}(\lambda), \quad \lambda = \Delta_1 q / \epsilon \tag{2.5}$$

satisfies ϵ-differential privacy (Dwork and Roth, 2014, Chapter 3).

Example 2.3. The discrete Gaussian distribution $\mathrm{DG}(\mu, \sigma^2)$ is defined by the probability mass function

$$f_{\mathrm{DG}}(x; \mu, \sigma^2) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sum_{j=-\infty}^{\infty} e^{-(j-\mu)^2/2\sigma^2}} \cdot \mathrm{I}(x \in \mathbb{Z}).$$

The associated discrete Gaussian mechanism

$$T_{\mathrm{DG}}(\boldsymbol{x}; q) = q(\boldsymbol{x}) + \boldsymbol{\xi}, \quad \xi_i \overset{\text{i.i.d.}}{\sim} \mathrm{DG}(0, \lambda^2), \quad \lambda = \Delta_1 q / \epsilon \tag{2.6}$$

satisfies $\epsilon^2/2$-differential privacy (Canonne et al., 2020, Theorem 4).

The addition of DP is considered separately from any post-processing, such as the constraints to non-negative integer values or hierarchical consistency utilized in the Census Bureau's TopDown Algorithm (Abowd et al., 2022). This post-processing can greatly affect the underlying DP distribution, as well as the associated properties. In this work we consider only the effects of adding DP noise, without any post-processing.

2.4 MCMC Sampling

From the perspective of a data analyst, a major benefit of differential privacy over other disclosure avoidance methodologies is that the noise-generating mechanism is often completely known; this includes any parameters. In principle, this means the noise mechanism can be included as a known source of uncertainty in a statistical analysis. However, selection of a noise mechanism may be concerned primarily with enforcing privacy criteria while maintaining a notion of utility of the underlying data; i.e., it is not necessarily chosen for the data analysts' convenience. Furthermore, future analyses involving the data may not be known to the agency while tabulations are being prepared for release.

We will compare two competing MCMC samplers under the Laplace mechanism mentioned in Section 2.3, and two other competing samplers under the Discrete Gaussian mechanism. In each case, one method uses the correct noise mechanism and one uses an approximation which is potentially more computationally efficient. Our objective is to ascertain whether the approximations produce nearly the correct inference, and whether the improvement in performance is worthwhile. The IID, REG, and CAR models will be considered under each of the two mechanisms. We will compare two Stan models (Carpenter et al., 2017) under a Laplace mechanism: one using the correct noise-generating Laplace mechanism, and one using a Gaussian approximation. Stan and other off-the-shelf modeling software is appealing for applied model development; once a desired model is correctly specified, the analyst need not derive, program, and debug a tailored sampling algorithm to apply it to the data. For the Discrete Gaussian noise mechanism, Stan currently does not support drawing discrete latent random variables; therefore, we will compare a Gibbs sampler making use of the correct Discrete Gaussian noise mechanism to a model using a continuous Gaussian approximation. The model with continuous Gaussian approximation is implemented using either Stan or a Gibbs sampler, depending on setting. The conditions necessitating the use of a Gibbs sampler are discussed in Section 6. Details for Gibbs samplers are given in Appendix A.

MCMC sampling will yield R draws of the parameter $\boldsymbol{\theta}$ and the latent process $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ from the posterior distribution $[\boldsymbol{Y}, \boldsymbol{\theta} \mid \boldsymbol{Z} = \boldsymbol{z}]$, denoted $\boldsymbol{\theta}^{(r)}$ and $\boldsymbol{Y}^{(r)}$, respectively, for $r = 1, \ldots, R$. A model-based prediction of Y_i is taken to be the mean $\widehat{Y}_i = \frac{1}{R}\sum_{r=1}^{R} Y_i^{(r)}$. A model-based 95% interval for Y_i is taken to be the 0.025 and 0.975 quantiles from the draws $Y_i^{(1)}, \ldots, Y_i^{(R)}$; the endpoints are respectively labeled $\widehat{Y}_i^L$ and $\widehat{Y}_i^U$.

Using the true noise-generating mechanism in the Laplace case, the data model assumes $\xi_i \overset{\text{i.i.d.}}{\sim} \text{Lap}(0, \lambda)$ for $i = 1, \ldots, n$. A competing data model is obtained using the approximation $\xi_i \sim \text{N}(0, \pi\lambda^2/2)$. Although other Gaussian approximations are possible, our choice of distribution is motivated by the following result.

Claim 2.4. Suppose $g(x; \nu, \lambda) = \frac{1}{2\lambda} e^{-|x-\nu|/\lambda}$ is a Laplace density with given parameters ν and λ. The Kullback-Leibler (KL) divergence between g and $f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$ is minimized when $\mu = \nu$ and $\sigma^2 = \pi\lambda^2/2$.

Proof. For two probability densities, f and g, with common support, the KL divergence from g to f is defined as $D_{\text{KL}}(f \mid g) = \text{E}_f(\log(f/g)) = \text{E}_f \log(f) - \text{E}_f \log(g)$. Easy calculations give $\text{E}_f \log(f) = -(1/2)\log(2\pi\sigma^2) - \text{E}_f(x-\mu)^2/(2\sigma^2)$ and $\log(g) = -\log(2\lambda) - |x-\nu|/\lambda$. Then

$$\begin{aligned} D_{\text{KL}}(f \mid g) &= \text{E}_f \log f - \text{E}_f \log g \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \text{E}_f \frac{(x-\mu)^2}{2\sigma^2} + \log(2\lambda) + \text{E}_f \frac{|x-\nu|}{\lambda} \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2} + \log(2\lambda) + \frac{1}{\lambda}\int |x-\nu| \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \end{aligned}$$

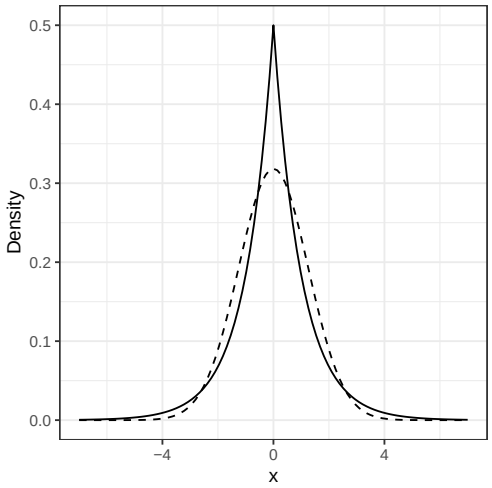


Figure 3: The Lap$(0, 1)$ density and Gaussian approximation N$(0, \pi/2)$ shown with solid and dashed lines, respectively.

Differentiating the integral

$$\int_{-\infty}^{\infty} |x-\nu| \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = -\int_{-\infty}^{\nu-\mu} (x+\mu-\nu)\, e^{-\frac{x^2}{2\sigma^2}} dx + \int_{\nu-\mu}^{\infty} (x+\mu-\nu)\, e^{-\frac{x^2}{2\sigma^2}} dx$$

with respect to μ and setting equal to zero gives $\mu = \nu$. We can therefore set $\mu = \nu = 0$ without loss of generality. Taking the second derivative, we have that this function is convex for $\mu = \nu = 0$. Then

$$\begin{aligned} D_{\mathrm{KL}}(f \mid g) &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2} + \log(2\lambda) + \frac{1}{\lambda}\,\mathrm{E}_f\,|x| \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2} + \log(2\lambda) + \frac{\sigma}{\lambda}\sqrt{\frac{2}{\pi}}. \end{aligned}$$

Differentiating the above with respect to σ and setting equal to zero gives the result. □

Figure 3 displays an example using Claim 2.4, with Lap$(0, 1)$ versus its Gaussian approximation N$(0, \pi/2)$.

3 Simulations with Laplace Mechanism

We conducted a simulation study to assess the suitability of the Gaussian approximation to the Laplace mechanism covered in Section 2.3 in a Gibbs sampler. Data were simulated according to data model (2.2) using IID, REG and CAR process models for $\boldsymbol{Y}$ in Sections 3.1, 3.2, and 3.3 respectively. Agency noise $\boldsymbol{\xi}$ variates were generated from the Laplace mechanism (2.5).

The primary goal was to predict $\boldsymbol{Y}$ using two MCMC samplers: one based on the correctly specified Laplace noise mechanism and one using the Gaussian approximation. We also compare uncertainty estimates and accuracy of credible intervals constructed using each sampler. A secondary goal was to evaluate accuracy of the estimates of unknown model parameters, specified for each scenario. Both samplers were implemented in Stan (Carpenter et al., 2017) through R (R Core Team, 2022) using the RStan package (Stan Development Team, 2022). When feasible, the default settings in Stan were used, which include a single chain of 2,000 HMC iterations, with the first 1,000 discarded as a burn-in, and a maximum tree-depth of 10. Adjustments needed to achieve good mixing and convergence as identified using appropriate diagnostics are noted in each of the respective sections.

For each setting of the simulation, $S = 100$ datasets were generated. Within the sth simulated dataset, a single chain of R draws was saved from each MCMC sampler indexed $r = 1, \ldots, R$. Let $Y_1^{(s)}, \ldots, Y_n^{(s)}$ denote the data in the sth dataset which represent the sensitive data that require protection. Let $\widehat{Y}_i^{(s)} = \frac{1}{R}\sum_{r=1}^{R} Y_i^{(r,s)}$ denote model-based predictions based on the R saved draws. Similarly, let $\widehat{Y}_i^{L(s)}$ and $\widehat{Y}_i^{U(s)}$ represent 0.025 and 0.975 quantiles from the draws $Y_i^{(1,s)}, \ldots, Y_i^{(R,s)}$. The following metrics are produced for the sth dataset for each sampler:

$$\text{RMSE}^{(s)} = \left[\frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_i^{(s)} - Y_i)^2\right]^{1/2}, \quad \text{MAE}^{(s)} = \frac{1}{n}\sum_{i=1}^{n}|\widehat{Y}_i^{(s)} - Y_i|, \quad \text{MRAE}^{(s)} = \frac{1}{n_+}\sum_{i:Y_i>0}\frac{|\widehat{Y}_i^{(s)} - Y_i|}{|Y_i|},$$

$$\text{MAX}^{(s)} = \bigvee_{i=1}^{n}|\widehat{Y}_i^{(s)} - Y_i|, \quad \text{COV}^{(s)} = \frac{1}{n}\sum_{i=1}^{n}\text{I}\left(\widehat{Y}_i^{L(s)} \le Y_i \le \widehat{Y}_i^{U(s)}\right), \quad \text{LEN}^{(s)} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Y}_i^{U(s)} - \widehat{Y}_i^{L(s)}\right),$$

where $n_+ = \sum_{i=1}^{n}\text{I}(y_i > 0)$ and $\bigvee_{i=1}^{n} x_i$ represents the maximum of $x_1, \ldots, x_n$. These metrics summarize root-mean squared error, bias, relative bias, maximum absolute error, and empirical coverage and length of an interval with 95% nominal coverage, respectively, over the n observations. We also consider "observed" versions of the metrics which use the observed differentially private measurements, Z_i, in place of model-based predictions:

$$\text{RMSE}^{(s)} = \left[\frac{1}{n}\sum_{i=1}^{n}(Z_i^{(s)} - Y_i)^2\right]^{1/2}, \quad \text{MAE}^{(s)} = \frac{1}{n}\sum_{i=1}^{n}|Z_i^{(s)} - Y_i|, \quad \text{MRAE}^{(s)} = \frac{1}{n_+}\sum_{i:Y_i>0}\frac{|Z_i^{(s)} - Y_i|}{|Y_i|},$$

$$\text{MAX}^{(s)} = \bigvee_{i=1}^{n}|Z_i^{(s)} - Y_i|, \quad \text{COV}^{(s)} = \frac{1}{n}\sum_{i=1}^{n}\text{I}\left(F_i^{-}(0.025 \mid \lambda_i) \le Z_i^{(s)} - Y_i \le F_i^{-}(0.975 \mid \lambda_i)\right),$$

$$\text{LEN}^{(s)} = \frac{1}{n}\sum_{i=1}^{n}\left[F_i^{-}(0.975 \mid \lambda_i) - F_i^{-}(0.025 \mid \lambda_i)\right] \equiv F_1^{-}(0.975 \mid \lambda_1) - F_1^{-}(0.025 \mid \lambda_1),$$

where $F_i^{-}(\gamma \mid \lambda_i)$ represents the γ quantile of the distribution of ξ_i. Simulation-level summaries for model-based predictions or observations are produced by averaging the S metrics using:

$$\text{RMSE} = \frac{1}{S}\sum_{s=1}^{S}\text{RMSE}^{(s)}, \quad \text{MAE} = \frac{1}{S}\sum_{s=1}^{S}\text{MAE}^{(s)}, \quad \text{MRAE} = \frac{1}{S}\sum_{s=1}^{S}\text{MRAE}^{(s)},$$

$$\text{MAX} = \frac{1}{S}\sum_{s=1}^{S}\text{MAX}^{(s)}, \quad \text{COV} = \frac{1}{S}\sum_{s=1}^{S}\text{COV}^{(s)}, \quad \text{LEN} = \frac{1}{S}\sum_{s=1}^{S}\text{LEN}^{(s)}.$$

To summarize draws of the parameters, let $\boldsymbol{\theta}^{(r,s)} = (\theta_1^{(r,s)}, \ldots, \theta_k^{(r,s)})$ denote rth saved draw of the parameter using the sth simulated dataset. Point estimates $\bar{\theta}_j^{(s)} = \frac{1}{R}\sum_{r=1}^{R}\theta_j^{(r,s)}$ are computed from the draws for the sth dataset, and summarized for the simulation using

$$\bar{\theta}_j = \frac{1}{S}\sum_{s=1}^{S}\bar{\theta}_j^{(s)} \quad \text{and} \quad \text{sd}(\theta_j) = \left[\frac{1}{S-1}\sum_{s=1}^{S}(\bar{\theta}_j^{(s)} - \bar{\theta}_j)^2\right]^{1/2}$$

for $j = 1, \ldots, k$. For each process model under consideration, we take privacy budget $\epsilon \in \{0.01, 0.1, 0.5\}$ and three settings of model parameters for a total of $3 \times 3 = 9$ settings. The values of ϵ correspond to $\lambda \equiv 1/\epsilon \in \{100, 10, 2\}$, which tend to generate large noise with $\lambda = 100$ and much smaller noise with $\lambda = 2$. Figure 1 suggests that the variance of census data increases with the mean, so we consider increasing the variance parameter along with the mean to mimic the behavior of the population counts for the selected detailed AIAN groups. The three settings of the model parameters are labeled Scenarios S1, S2, and S3 for each form of the process model (IID, REG, and CAR) and their exact meanings are specified in the respective subsections.

3.1 Independent and Identically Distributed

Under the IID process model, data were generated with sample size $n = 1{,}000$. We consider three settings of the parameters (μ, σ^2): Scenario S1 uses $\mu = 1$ and $\sigma^2 = 1$, S2 uses $\mu = 2$ and $\sigma^2 = 2$, and S3 uses $\mu = 3$ and $\sigma^2 = 3$. The Laplace sampler produced a chain of 3,000 HMC iterations, with the first 1,000 iterations discarded as a burn-in, yielding $R = 2{,}000$ saved draws, while the Gaussian sampler produced a chain of 2,000 HMC iterations, with the first 1,000 discarded as a burn-in. We found that Stan is very sensitive to the choice of starting values, as the chain moves around the parameter space very slowly. Initial values

$$\boldsymbol{Y}^{(0,s)} = \boldsymbol{W}^{(s)}, \quad \mu^{(0,s)} = \frac{1}{n}\sum_{i=1}^{n} W_i^{(s)}, \quad \text{and} \quad \sigma^{2(0,s)} = \frac{1}{n-1}\sum_{i=1}^{n}\left[W_i^{(s)} - \mu^{(0,s)}\right]^2$$

helped to achieve good HMC chains, where $W_i^{(s)} = \log\{\max(Z_i^{(s)}, 1)\}$ and $\boldsymbol{Y}^{(r,s)}$ represents the rth draw of the latent random variable $\boldsymbol{Y}^{(s)}$. Here we have assumed Stan's default improper prior for μ by leaving the distribution unspecified, and assumed a prior $\sigma^2 \sim \text{Gamma}(10, 10)$.

The samplers for the entire simulation took about 46.74 hours to complete: 43.38 hours for the Laplace sampler and 3.37 hours for the Gaussian sampler. Average sampler run times for each scenario are displayed in Table 13. Increasing (μ, σ^2) for each fixed ϵ appears to only have had a mild effect on the runtimes of the Gaussian model, but caused a more severe increase in run times of the Laplace model.

In Scenario S1, where the Laplace mechanism produced the most noise and the "signal" from the process model was lowest, traceplots (not shown) provided evidence that the chains from both samplers did not mix well. The Gaussian approximation model appeared to fare better than the Laplace model however. Here it appears that, even though the distribution of the noise is known, characteristics due to the process model are drowned out by the amount of noise. In general, the mixing of the chains improves as ϵ or (μ, σ^2) are increased, corresponding to a larger signal to noise ratio.

Tables 1 and 2 summarize predictions and parameter estimates respectively, reporting the metrics described in the beginning of Section 3. Ignoring the extremely noisy and low signal Scenario S1, the Laplace model produced more accurate estimates of μ and σ^2 than the Gaussian approximation model. However, when ϵ is reduced or (μ, σ^2) is increased, the two samplers produce very similar results for the estimates. The prediction results in those scenarios are comparable as well. A notable difference is that the length LEN of the intervals is wider from the Laplace model, as we might expect, so that the coverage COV of the intervals from the Gaussian approximation model is slightly lower than the nominal 95% level.

3.2 Regression

Under the REG process model with regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1)$, Scenario S1 assumes $\beta_1 = 0.5$ and $\sigma^2 = 1$, S2 assumes $\beta_1 = 1.5$ and $\sigma^2 = 2$, and S3 assumes $\beta_1 = 2.5$ and $\sigma^2 = 3$; $\beta_0 = 0.5$ remains fixed. Using sample size $n = 1{,}000$, the design matrix $\boldsymbol{X}$ was constructed from two columns: an intercept with all ones, and a second column whose ith row is generated by $x_i \sim \text{N}(1, 1)$ and fixed throughout the simulation. Thus, we mimic the settings used in the IID process model with $\mu = \boldsymbol{X}\boldsymbol{\beta}$. The Laplace sampler was run to produce a chain of length 3,000, with the first 1,000 iterations discarded as burn-in, yielding $R = 2{,}000$ saved draws, while the Gaussian sampler was run to produce a chain length of 2,000, with the first 1,000 iterations discarded as burn-in.

Similar to the IID setting, initial values for Stan were taken to be

$$\boldsymbol{Y}^{(0,s)} = \boldsymbol{W}^{(s)}, \quad \boldsymbol{\beta}^{(0,s)} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{W}^{(s)}, \quad \text{and} \quad \sigma^{2(0,s)} = \frac{1}{n-1}\sum_{i=1}^{n}\left[W_i^{(s)} - \bar{W}^{(s)}\right]^2$$

where $\bar{W}^{(s)} = \frac{1}{n}\sum_{i=1}^{n} W_i^{(s)}$ and $W_i^{(s)} = \log\{\max(Z_i^{(s)}, 1)\}$. Here $\boldsymbol{\beta}^{(0,s)}$ has been taken to be the ordinary least squares estimator from the regression of $\boldsymbol{W}^{(s)}$ on $\boldsymbol{X}$. For both the Laplace and Gaussian approximation model, we assumed priors $\sigma^2 \sim \text{Gamma}(10, 10)$ and $\boldsymbol{\beta} \sim \text{N}(\boldsymbol{0}, 25\boldsymbol{I})$.

The samplers for the entire simulation took 194 hours to complete, with average sampler run times for each scenario displayed in Table 13. As in the IID case, we see that fitting the Laplace model was much slower than fitting the Gaussian approximation model. On average, across all simulations and scenarios, the Laplace sampler took 13 times as long as compared to the Gaussian sampler. For the Laplace model, we encountered poor mixing and slow convergence with the default settings in Stan. We increased the maximum tree depth to 13, and also increased the maximum number of iterations to 3,000 to address these convergence issues, though at the cost of longer running times. We did not encounter these issues with the sampler for the Gaussian model using the default arguments. As in the IID setting, chains in Scenario S1 showed evidence of inadequate mixing for both samplers; we again set it aside to focus on the other scenarios.

Tables 3 and 4 give predictions and parameter estimates, respectively. In general, we found that the model-based estimates had higher accuracy as compared to the noisy direct estimates. As expected, this was especially true for the scenarios with higher noise infusion. The results of the Laplace and Gaussian approximation model were largely similar. One exception was the case of the high DP noise, and small μ and σ^2, in which case the Laplace model had much less error and bias. Overall, the Gaussian approximation tended to have better coverage of the 95% interval, with similar interval lengths.

3.3 Conditional Autoregression

Under the CAR process model, the levels of $\boldsymbol{\beta}$ and setup of $\boldsymbol{X}$ were selected in the same manner as in Section 3.2 for the REG model. We used a 10×10 square grid to construct the adjacency matrix $\boldsymbol{A}$ discussed in Section 2.1, with one observation per area so that the sample size is $n = 100$. The lattice structure and size were selected to illustrate a simple case of the CAR structure, while also maintaining reasonable runtimes for the simulation. Both models were fit with 3,000 HMC iterations, with the first 1,000 iterations discarded as a burn-in, yielding $R = 2{,}000$ saved draws. The parameter ρ for CAR was fixed at 0.95 throughout, and assumed known in the model. Initial values for Stan were formed in the same way as the REG setting. For both the Laplace and Gaussian approximation model, we assumed priors $\sigma^2 \sim \text{Gamma}(10, 10)$ and $\boldsymbol{\beta} \sim \text{N}(\mathbf{0}, 25\boldsymbol{I})$.

MCMC sampling for the entire simulation took 904 hours (≈ 37.67 days) to complete, with average times per scenario given in Table 13. The Laplace sampler was much slower than the Gaussian sampler, with the Laplace model requiring on average 6.1 times as much time as the Gaussian model. We encountered similar mixing and convergence issues as in the REG case, primarily with the Laplace model, though to a lesser extent with the Gaussian approximation. Increasing the maximum iterations to 3,000, and increasing the maximum tree depth to 13 again helped to address these issues.

Tables 5 and 6 give predictions and parameter estimates, respectively. The results here largely mimicked the findings in the REG case. The model based estimates far outperformed the noisy DP measurements. The Gaussian approximation performed similarly to the Laplace, even in the case of high noise and low μ and σ^2. Coverage rates as well as interval length for the 95% interval were similar between the two model-based approaches.

4 Simulations with Discrete Gaussian Mechanism

This section repeats Section 3 but with Discrete Gaussian as the noise mechanism. Data were simulated according to the IID, REG, and CAR process models for $\boldsymbol{Y}$ in Sections 4.1, 4.2, and 4.3, respectively, and agency noise $\boldsymbol{\xi}$ variates were generated from the Discrete Gaussian mechanism (2.6). The objective now is to compare predictions for $\boldsymbol{Y}$ using two MCMC samplers: a Gibbs sampler using the correctly specified Discrete Gaussian noise mechanism with $\xi_i \overset{\text{i.i.d.}}{\sim} \text{DG}(0, \lambda^2)$ in the data model and Stan using a continuous Gaussian distribution $\xi_i \overset{\text{i.i.d.}}{\sim} \text{N}(0, \lambda^2)$ in its place. The simulation setup and metrics are largely the same as in Section 3, but some of the information is repeated in the following subsections for readability.

4.1 Independent and Identically Distributed

Under the IID process model, data are generated with sample size $n = 1{,}000$. We consider three settings of the parameters (μ, σ^2): Scenario S1 uses $\mu = 1$ and $\sigma^2 = 1$, S2 uses $\mu = 2$ and $\sigma^2 = 2$, and S3 uses $\mu = 3$ and $\sigma^2 = 3$. Based on initial diagnostics, both samplers are run to produce chains of 3,000 iterations, with the first 2,000 discarded as a burn-in, yielding $R = 1{,}000$ saved draws. Initial values are taken to be

$$\boldsymbol{Y}^{(0,s)} = \boldsymbol{W}^{(s)}, \quad \mu^{(0,s)} = \frac{1}{n}\sum_{i=1}^{n} W_i^{(s)}, \quad \text{and} \quad \sigma^{2(0,s)} = \frac{1}{n-1}\sum_{i=1}^{n}\left[W_i^{(s)} - \mu^{(0,s)}\right]^2,$$

where $W_i^{(s)} = \log\{\max(Z_i^{(s)}, 1)\}$ and $\boldsymbol{Y}^{(r,s)}$ represents the rth draw of the latent random variable $\boldsymbol{Y}^{(s)}$. The prior is taken to be $\mu \sim \mathrm{N}(0, 25)$ and $\sigma^2 \sim \mathrm{IG}(2, 10)$.

Total MCMC sampling time over the whole simulation took 253.76 hours to complete. Average sampler run times for each scenario are displayed in Table 14. Stan run times are generally faster than the Gibbs sampler, except in Scenario S3 with $\epsilon = 0.5$ (low noise and high signal) where it slows down very noticeably.

Tables 1 and 2 summarize predictions and parameter estimates respectively. In Scenario S1 with $\epsilon = 0.01$, both IID models show signs of poor performance: the Gaussian model has a COV of 0.39 while LEN for the DG model appears to be extremely wide. Metrics appear to be improved in other IID models aside from Scenario S3 with $\epsilon = 0.01$, where the RMSE of Gaussian is much larger than for DG; in this high signal and high noise setting, traceplots from Stan showed evidence of slow mixing for several of the simulated datasets which appear to be the cause. This effects of this can also be seen in Table 2 as $\mathrm{sd}(\mu)$ and $\mathrm{sd}(\sigma^2)$ increase from $\epsilon = 0.1$ to $\epsilon = 0.5$ under Scenario S3.

4.2 Regression

Under the REG process model, values for regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1)$ are taken to be $(0.5, 0.5)$ for Scenario S1, $(0.5, 1.5)$ for S2, and $(0.5, 2.5)$ for S3. Using sample size $n = 1{,}000$, the design matrix $\boldsymbol{X}$ is constructed from two columns, an intercept with all ones and a second column whose ith row is generated by $x_i \sim \mathrm{N}(1, 1)$, and fixed throughout the simulation. As in the IID case, samplers produced chains of length 3,000, with the first 2,000 iterations discarded as burn-in, yielding $R = 1{,}000$ saved draws. Initial values are taken to be

$$\boldsymbol{Y}^{(0,s)} = \boldsymbol{W}^{(s)}, \quad \boldsymbol{\beta}^{(0,s)} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{W}^{(s)}, \quad \text{and} \quad \sigma^{2(0,s)} = \frac{1}{n-1}\sum_{i=1}^{n}\left[W_i^{(s)} - \bar{W}^{(s)}\right]^2,$$

where $\bar{W}^{(s)} = \frac{1}{n}\sum_{i=1}^{n} W_i^{(s)}$ and $W_i^{(s)} = \log\{\max(Z_i^{(s)}, 1)\}$. Here $\boldsymbol{\beta}^{(0,s)}$ is taken to be the ordinary least squares estimator from the regression of $\boldsymbol{W}^{(s)}$ on $\boldsymbol{X}$. We assume priors $\sigma^2 \sim \mathrm{IG}(2, 10)$ and $\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{0}, 25\boldsymbol{I})$.

Total MCMC sampling time over the whole simulation took 267.58 hours to complete. Average sampler run times for each scenario are displayed in Table 14. Here, Gibbs takes substantially longer than Stan, with Stan run times increasing under Scenario S3 as ϵ increases.

Tables 9 and 10 summarize predictions and parameter estimates respectively. As in the IID case, COV and LEN results for Scenario S1 with $\epsilon = 0.01$ indicate problems for both samplers. Here, trace plots (not shown) provide evidence poor mixing in both cases. Results in remaining cases appear to be comparable with no indication of problems; the most obvious exception being a low COV for Scenario S3 with $\epsilon = 0.5$, which appears to experience slow mixing under several of the generated datasets.

4.3 Conditional Autoregression

Under the CAR process model, the levels of $\boldsymbol{\beta}$ and setup of $\boldsymbol{X}$ are selected in the same manner as in Section 4.2 for the REG model. We use a 10×10 lattice to construct the adjacency matrix $\boldsymbol{A}$ discussed in Section 2.1, with one observation per area so that the sample size is $n = 100$. Samplers are run to produce chains of length 3,000, with the first 1,000 draws discarded as burn-in, yielding $R = 2{,}000$ saved draws. The

parameter ρ for CAR is fixed at 0.95 throughout and assumed known in the model. A similar analysis could be conducted without an estimate of ρ using an intrinsic conditional autoregressive (ICAR) model (Besag and Kooperberg, 1995). Initial values for Stan are formed in the same way as the REG setting. The prior is taken to be $\sigma^2 \sim \text{IG}(2, 10)$ and $\boldsymbol{\beta} \sim \text{N}(\mathbf{0}, 25\boldsymbol{I})$.

Total MCMC sampling time over the whole simulation took 113.70 hours to complete. Average sampler run times for each scenario are displayed in Table 14. Here, Stan tends to take longer than the Gibbs sampler, especially with the CAR latent process model.

Tables 11 and 12 give predictions and parameter estimates, respectively. As in the IID and REG cases, but to a smaller degree, COV and LEN appear to suffer under Scenario S1 with $\epsilon = 0.01$. Remaining results show no evidence of mixing issues or poor performance. Furthermore, results from the two samplers appear to correspond well. Table 12 also shows good agreement between the two samplers. However, due to the confounding phenomenon under the CAR model, coefficient values themselves may not be comparable to the true values as the inclusion of a spatially correlated random effect can drastically affect estimation of fixed-effects coefficients (e.g., Hughes and Haran, 2013; Hodges and Reich, 2010; Bradley et al., 2015).

5 Empirical Simulation Study with Laplace Mechanism

This section presents the results of an empirical simulation study using 2010 census counts of the number of persons in the 77 counties in Oklahoma for each of the AIAN race groups Cherokee, Choctaw, and Sioux. This provides a realistic setting to compare models using the exact noise mechanism with models using an approximating Gaussian data model, and to understand the benefit of including regression terms and spatially correlated random effects in a BHM fit to real data. This study is also used to quantify the gains in precision using model-based methods of increasing complexity to estimate counts of persons compared to directly using noisy, differentially private measurements, and to understand the computational tradeoffs of using the exact, but difficult to fit Laplace data model, vs. using the approximating Gaussian data model.

In this study, the released 2010 census counts are treated as sensitive data which are protected by a Laplace noise mechanism before release (without any post-processing). Each of the three AIAN groups is taken individually and privacy protected using three levels of privacy protection, $\epsilon \in \{0.01, 0.1, 0.5\}$, giving a total of 9 simulation levels. These values of ϵ were chosen to roughly represent very high, medium, and low levels of privacy protection/noise. For each level, datasets $\boldsymbol{Z}^{(s)} = (Z_1^{(s)}, \ldots, Z_n^{(s)})$ for $s = 1, \ldots, S = 100$ are generated by adding Laplace noise to the 2010 counts $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ for the corresponding AIAN group; i.e.

$$Z_i^{(s)} = Y_i + \xi_i^{(s)}, \quad \xi_i^{(s)} \overset{\text{i.i.d.}}{\sim} \text{Lap}(0, 1/\epsilon). \tag{5.1}$$

Here, $Z_i^{(s)}$ represents the noisy, differentially private measurement of the true, unobservable 2010 census count, Y_i for the ith county. With 100 datasets generated for each of the 9 simulation levels, a total of 900 datasets were generated.

The IID and REG models were fit to each dataset $\boldsymbol{Z}^{(s)}$ in two ways: first using the correct Laplace noise mechanism in the data model, then replacing the Laplace mechanism with our Gaussian approximation. For the REG model, the design matrix $\boldsymbol{X}$ consists of three columns: an intercept, the logarithm of corresponding 2000 census count incremented by one, and the logarithm of the corresponding ACS 2009 count incremented by one. The logarithm of the 2000 census and 2009 ACS counts, summarized in Figure 1, are regarded as non-sensitive, publicly available data sources that can be used to model 2010 census counts.

We also fit a model which includes a spatially-correlated random effect, but rather than using the CAR model given in Section 3.3, we used an ICAR model (Besag and Kooperberg, 1995). Using an ICAR distribution on the random effects avoids the difficulties of estimating the parameter ρ in the CAR model presented in 2.2. The ICAR prior is also advantageous in that we found it results in much faster fitting times in Stan than when using a CAR prior. The model, with an approximating Gaussian noise distribution, is

then

$$
\begin{aligned}
Z_i &= Y_i + \xi_i, \quad \xi_i \sim \mathrm{N}\left(0, \pi\lambda_i^2/2\right), \\
\log Y_i &= \boldsymbol{x}_i^\top \boldsymbol{\beta} + \eta_i + \gamma_i, \quad \gamma_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \phi^2), \\
\boldsymbol{\eta} &= (\eta_1, \ldots, \eta_n)^\top \,|\sigma^2 \sim \mathrm{ICAR}(\sigma^2),
\end{aligned}
$$

where $\mathrm{ICAR}(\sigma^2)$ denotes the ICAR distribution parameterized by the single variance component, σ^2. The ICAR distribution can be characterized by its full conditional distributions. Let $\mathcal{N}_i$ index the set of geographies which neighbor area i, and let n_i be the number of elements in $\mathcal{N}_i$. Then

$$
\eta_i \mid \boldsymbol{\eta}_{-i}, \sigma^2 \sim \mathrm{N}\left(\frac{1}{n_i}\sum_{j \in \mathcal{N}_i} \eta_j, \frac{\sigma^2}{n_i}\right),
$$

where $\boldsymbol{\eta}_{-i}$ represents the vector $\boldsymbol{\eta}$, with the ith element removed. To avoid identifiability issues, we impose the usual constraint that $\sum_i \eta_i = 0$. Random errors γ_i representing fine scale variability were added to improve mixing in Stan. The same initial values as the REG model were used, in addition to $\boldsymbol{\eta} = \mathbf{0}$. For the priors, we used $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I})$ and $\sigma^2 \sim \mathrm{Gamma}(a_\sigma, b_\sigma)$ with $\sigma_\beta = 5$, $a_\sigma = 10$, $b_\sigma = 10$.

Mathematically, this model is a fairly straightforward extension of the REG model, but the computational challenges of fitting this model increased dramatically over fitting the REG model. The time increased substantially fitting the ICAR model over the other models using equivalent settings in Stan. Also, keeping the Stan settings the same resulted in poor model fit, as evidenced by examination of the traceplots and analysis of the predicted values. One issue may be that, given the strength of spatial dependence exhibited, the number of areas used (i.e., the 77 counties in Oklahoma) is too small to fit a spatial model. See Janicki et al. (2022) for similar issues determining the number of geographic regions to use in a spatial areal model, and discussion on tradeoffs with increasing or decreasing the number of areas. We therefore expanded the number of geographies to include counties from Oklahoma's six neighboring states, Colorado, Kansas, Missouri, Arkansas, Texas, and New Mexico, for a total of 723 counties. Based on initial investigations, we also increased the number of HMC iterations to 3,000, with the first 1,500 discarded as warmup, and increased the maximum tree depth tuning parameter to 15 to improve convergence. Adjusting the tuning parameters to these settings resulted in much better model fit, as evidenced by examination of the traceplots and comparison of model-based predictions to the true values. The tradeoff with choosing these HMC settings is the massive increase in computational time, as can be seen in Table 15. Due to these challenges, we only present results for an empirical simulation study using the data set consisting of counts of Cherokee persons by county.

As challenging as it was to fit this model with a Gaussian noise approximation, fitting the model using the true Laplace model was greater still. The time to fit the Gaussian model to a single data set, in this example, was over 200 minutes on average across all settings, as can be seen in Table 15. The time to fit the Laplace model to a single data set, using the same tuning parameters as with the Gaussian model, was over a day, and still convergence was not achieved. The lack of convergence was obvious, and could be observed by examining the traceplots of the predicted values of the true counts (not shown). Despite using reasonable starting values and increasing the number of HMC iterations, we could not achieve convergence using the exact Laplace data model in under 24 hours in Stan.

For each of the three models, IID, REG, and ICAR, we used initial values in the HMC sampler in Stan of $Y_i^* = \max\{Z_i, 1\}$ for Y_i, $i = 1, \ldots, n$. The initial value used for σ^2 was the variance of $a_1 = \log Y_1^*, \ldots, a_n = \log Y_n^*$. For models REG and ICAR, the initial value used for $\boldsymbol{\beta}$ was $\boldsymbol{\beta}^* = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{a}$ where $\boldsymbol{a} = (a_1, \ldots, a_n)$. The adjacency matrix $\boldsymbol{A}$ for the ICAR model was computed from the geography of Oklahoma counties as well as counties in surrounding states. For each of the three models, draws $Y_i^{(r,s)}$ for $r = 1, \ldots, R$ were saved so that metrics described in Section 3 could be computed. HMC for all models was carried out using Stan. We found that the Stan tuning parameters and the number of HMC iterations to achieve convergence needed to be chosen separately for each model. In total, the samplers for the IID model

took 4.89 hours to run, while the samplers for the REG model took 3.07 hours. The Stan samplers for the three Cherokee scenarios with the Gaussian ICAR model took 1022.23 hours to run.

Table 17 gives an overall summary of the performance of the predictions for each scenario and for each detailed race group. Table 17 shows the root mean squared error (RMSE), mean absolute error (MAE), mean relative absolute error (MRAE), maximum deviation (MAX), coverage rate (COV), and interval length (LEN), averaged over all 100 simulations for each scenario and for each detailed AIAN race group. We note while the same noise levels are used with the Cherokee, Choctaw, and Sioux data sets, that comparisons across these detailed AIAN race groups are not necessarily meaningful, as the predictions depend not only on the noise levels, but also characteristics of the underlying true values as well as the predictors. Differences in the summary statistics across columns with common noise levels in the Observed block of Table 17 are due to simulation noise.

Comparing results within each column of Table 17 we see that model-based predictions generally outperform the noisy measurements under each of the given metrics, regardless of the amount of noise added, and regardless of whether the true Laplace data model or the approximating Gaussian data model is used. The RMSE decreases substantially in each scenario using the IID model, compared to the noisy measurements, especially with the Sioux data set. The predictions using the IID model are also well calibrated, in the sense that the coverage rates of the credible intervals are close to the nominal 95% level. Using the correctly-specified Laplace model gives better results than the approximating Gaussian data model, particularly with uncertainty estimates and credible intervals, but the point estimates are generally very comparable, indicating that the approximate Gaussian model gives good results.

Adding covariate information to the model, as can be seen in the REG blocks of Table 17, further increases the precision of predictions, particularly in the very high noise case, $\epsilon = 0.01$. Again, the Laplace model and Gaussian model give very similar results. We do, however, see a degradation in performance of the credible intervals, as coverage rates tend to be in the 80s. This could be due to not accounting for the uncertainty of the ACS 5-year estimates, the specification of the hyperparameters or prior distributions, or the tuning of the HMC algorithm.

The ICAR model, perhaps unsurprisingly, gives the best overall results. Accounting for spatial correlation using an ICAR prior on the random effects both further improves the precision of the point predictions in the $\epsilon = 0.1$ and 0.5 scenarios, and also corrects the coverage rates in all scenarios. The worse performance of the point estimates in the $\epsilon = 0.01$ scenario may be more due to issues of fit using the REG model with very high noise levels, rather than with the ICAR model.

Overall, we see that the Gaussian approximation to the data model gives very similar performance to using the exact Laplacian model in nearly every scenario, and can be used in situations where a Laplacian model is not feasible, such as when accounting for spatial correlation in the model. The greatest modeling benefits are when there are very high levels of added noise. With very low levels of noise, the benefits are far more modest, although there does seem to be some improvement in precision. Regardless of the amount of added noise or the degree of added precision of the point estimates, the ability to fit a model to the observed, noisy data, is important, as it allows for prediction of out of sample quantities, as in change of support and downscaling problems, which is an important area of future research.

6 Empirical Simulation Study with Discrete Gaussian Mechanism

This section presents a similar study as in Section 5 based on the same AIAN data, outcomes, and covariates. However, the observations are now prepared with a Discrete Gaussian mechanism so that

$$Z_i^{(s)} = Y_i + \xi_i^{(s)}, \quad \xi_i^{(s)} \overset{\text{i.i.d.}}{\sim} \text{DG}(0, 1/\epsilon) \tag{6.1}$$

for county $i = 1, \ldots, n$ within the sth simulated dataset $s = 1, \ldots, S$. Here, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ are taken to be the released 2010 census counts for either Cherokee, Choctaw, or Sioux groups, and three levels of noise $\epsilon \in \{0.01, 0.1, 0.5\}$ are considered. There were $S = 100$ simulated datasets generated within each of the nine simulation settings.

Three pairs of models were considered to fit the generated data based on the IID, REG, and CAR models. For each of the IID, REG, and CAR model types, one variant was fit using the true DG noise mechanism in the data model and a second variant was fit with an associated Gaussian distribution in place of the DG mechanism. We refer to these models using the labels DG IID, Gauss IID, DG REG, Gauss REG, DG CAR, and Gauss CAR. The Gauss IID and Gauss REG models were fit with Stan as in Section 5.

Because Stan does not directly support discrete latent variables, the Gibbs sampling algorithm discussed in Appendix A was used to fit the DG IID, DG REG, and DG CAR models. Furthermore, initial attempts to produce suitable chains for the Gaussian CAR model using our Stan code were unsuccessful; therefore, a Gibbs sampling approach was also used in this setting. Remark A.2 demonstrates that the obtained Gibbs sampler cannot estimate the intercept parameter under an ICAR assumption; therefore, the CAR setting uses a proper CAR structure with fixed dependency parameter $\rho = 0.95$. As in Section 4, all simulations were fit by generating a chain of 3,000 iterations and discarding the first 2,000 as a burn-in sample. A common conjugate prior of $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 25\boldsymbol{I})$ and $\sigma^2 \sim \mathrm{IG}(2, 10)$ was used in all cases. A maximum tree depth of 12 was used for Stan with the Gaussian REG model. In total, the samplers for the IID case took 155.92 hours, the samplers for the REG case took 268.13 hours, while the CAR case took 350.32 hours.

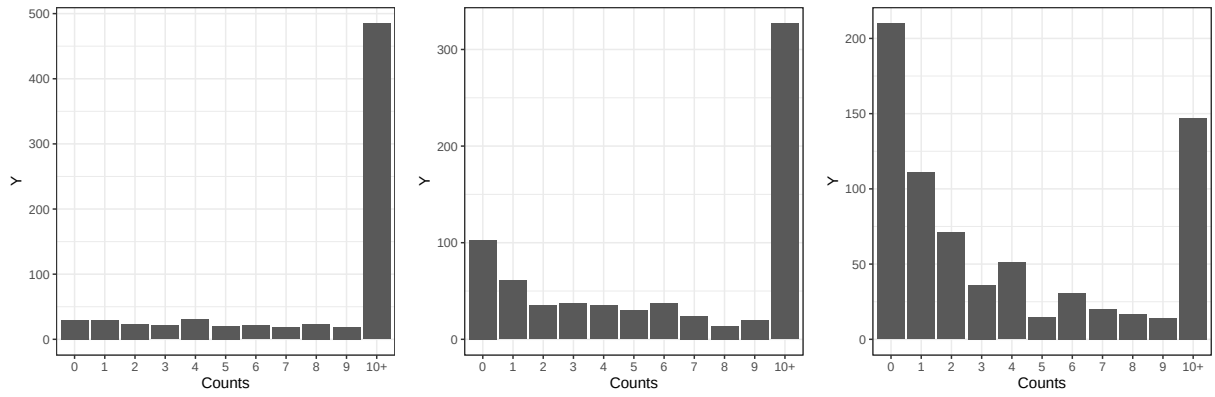
Table 18 summarizes prediction results from the simulation and Table 16 displays the associated runtimes for reference. As anticipated, the RMSE, MAE, and MRAE tend to improve from Observed to IID, REG, and CAR, as models become more expressive. DG and Gaussian results tend to be similar in most cases. Focusing on COV, it appears that models for the Cherokee data attain closest to the nominal 95% coverage level, while the Sioux data tend to be furthest. Some cases such as DG CAR with Sioux data have poor coverage and a very narrow LEN. In these cases, neither the DG nor Gaussian models appeared to mix well under the large noise setting $\epsilon = 0.01$, but improved as ϵ was increased to 0.5. The large noise may be especially problematic with the Sioux data because of the large number of zero counts; this can be seen in Figure 4.

A closer look at the Gibbs sampler reveals some issues that may be contributing to the low coverage. With integer-valued responses and DG noise in the data model, the choice of transformation $f(x) = \log x$ precludes drawing integer values of zero from the conditional $[y_i \mid \boldsymbol{y}_{-i}, \boldsymbol{\theta}]$; see Remark A.1. This affects the COV metric when there are many zeros in the data. Future work with integer-valued response and DP noise would benefit from the transformation $f(x) = \log(x + 1)$ instead. Also, with DG noise in the data model, the Gibbs sampler was seen to get "stuck" in regions where only very small moves were possible. Here the elements of $\boldsymbol{y}$ became unlikely to change by a full integer; trace plots of $\boldsymbol{\beta}$ and σ^2 appear to show good mixing, but their exploration is constrained due to the lack of movement of $\boldsymbol{y}$. Under similar circumstances with a Gaussian data model, all components of the chain are able to move, and evidence of poor mixing for $\boldsymbol{\beta}$ and σ^2 can more readily be seen. The underlying mixing issue might suggest that the Lognormal model may not be ideal in this setting.

7 Conclusions

We have investigated several differentially private noise mechanisms in the data model within a hierarchical Bayesian model. Specifically, Laplace and Discrete Gaussian mechanisms were compared to respective Gaussian approximations in a series of simulations. We first considered data generated from a Gaussian distribution conditional on a lognormally distributed process model, then proceeded to empirical simulations with noise mechanisms applied to selected counts from the 2010 census.

Unsurprisingly, in cases where the noise mechanism has a large magnitude to drown out the signal from the latent process, models have more difficulty capturing the process. This manifests in poor mixing of MCMC chains which is evident in diagnostics such as trace plots and in the summarized results. Evidence of poor mixing from Stan is also seen with small magnitude noise and large signal, but is not seen in Gibbs samplers. For empirical simulations, the featured Lognormal model works well when small counts are not too prevalent and noise magnitude is not too large. Gaussian approximations appear to give very similar results to models using the exact Laplace or Discrete Gaussian noise mechanisms, provided that the noise magnitude is not too large; in this case, use of the exact noise mechanism is preferable. However, runtimes in



(a) Cherokee. (b) Choctaw. (c) Sioux.

Figure 4: Distribution of counts from the 2010 census for the three AIAN groups in Oklahoma and neighboring states.

Stan appear to suffer when outcomes are assumed to be sums of Lognormal and Laplace random variables. Note that results obtained from Stan which were less ideal may be a symptom of our programming rather than the capabilities of Stan.

Modeling of differentially private data will continue to be of interest as statistical agencies adopt it for releases. Our study assumed fairly standard latent process models and was limited to a simplified dataset extracted from a vastly more complicated release. The recent and future data releases feature many levels of geography, many other race and ethnicity groups, and a number of tabulations which are subsetted to particular groups in the overall population and crossed with a number of other factors of interest. Future work will consider more holistic models to capture dependencies among tabulations within a release and between multiple data products.

References

John Abowd, Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, Brett Moran, William Sexton, Matthew Spence, and Pavel Zhuravlev. The 2020 Census disclosure avoidance system TopDown algorithm. *Harvard Data Science Review*, (Special Issue 2), 2022. doi: https://doi.org/10.1162/99608f92.529e3cb9. URL https://hdsr.mitpress.mit.edu/pub/7evz361i.

John M. Abowd. The U. S. Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*, New York, NY, USA, 2018. Association for Computing Machinery. URL https://doi.org/10.1145/3219819.3226070.

John M. Abowd, Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, Brett Moran, William Sexton, Matthew Spence, and Pavel Zhuravlev. The Census TopDown algorithm, 2021. Manuscript submitted to Harvard Data Science Review.

Julian Besag and Charles Kooperberg. On conditional and intrinsic autoregression. *Biometrika*, 82(4): 733–746, 1995. ISSN 00063444. URL http://www.jstor.org/stable/2337341.

Jonathan R. Bradley, Scott H. Holan, and Christopher K. Wikle. Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. *The Annals of Applied Statistics*, 9(4):1761–1791, 2015. URL https://www.jstor.org/stable/43826444.

Clément L. Canonne, Gautam Kamath, and Thomas Steinke. The discrete Gaussian for differential privacy. *CoRR*, abs/2004.00010, 2020. URL https://arxiv.org/abs/2004.00010.

Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32, 2017. ISSN 1548-7660. doi: 10.18637/jss.v076.i01. URL https://www.jstatsoft.org/v076/i01.

N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. Wiley, 2015. URL https://books.google.com/books?id=KsDdCgAAQBAJ.

Noel A.C. Cressie. *Statistics for Spatial Data*. Wiley-Interscience, revised edition, 1993.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9:211–407, 2014.

James S. Hodges and Brian J. Reich. Adding spatially-correlated errors can mess up the fixed effects you love. *The American Statistician*, 64(4):325–344, 2010. doi: https://doi.org/10.1198/tast.2010.10052.

John Hughes and Murali Haran. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):139–159, 2013. doi: https://doi.org/10.1111/j.1467-9868.2012.01041.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2012.01041.x.

Ryan Janicki, Andrew M. Raim, Scott H. Holan, and Jerry J. Maples. Bayesian nonparametric multivariate spatial mixture mixed effects models with application to American Community Survey special tabulations. *The Annals of Applied Statistics*, 16(1):144–168, 2022. doi: https://doi.org/10.1214/21-AOAS1494. URL https://doi.org/10.1214/21-AOAS1494.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL https://www.R-project.org/.

Andrew M. Raim. Direct sampling in Bayesian regression models with additive disclosure avoidance noise. Research Report Series: Statistics #2021-01, Center for Statistical Research and Methodology, U.S. Census Bureau, 2021. URL https://www.census.gov/library/working-papers/2021/adrm/RRS2021-01.html.

Stan Development Team. RStan: the R interface to Stan, 2022. URL https://mc-stan.org/. R package version 2.21.5.

Kyle Walker and Matt Herman. *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*, 2022. URL https://CRAN.R-project.org/package=tidycensus. R package version 1.2.

Stephen G. Walker, Purushottam W. Laud, Daniel Zantedeschi, and Paul Damien. Direct sampling. *Journal of Computational and Graphical Statistics*, 20(3):692–713, 2011. doi: 10.1198/jcgs.2011.09090. URL https://doi.org/10.1198/jcgs.2011.09090.

A Gibbs Samplers with Direct Sampling for Noise

Define the following notation.

- Let $\log(\boldsymbol{y}) = (\log y_1, \ldots, \log y_n)$ denote elementwise application of the logarithm.
- A random variable $\boldsymbol{W} \in \mathbb{R}^n$ has multivariate normal distribution $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with density $f_{\mathrm{N}}(\boldsymbol{w} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \exp\{-\frac{1}{2}(\boldsymbol{w}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{w}-\boldsymbol{\mu})\}$.

- A random variable $\boldsymbol{W} \in \mathbb{R}^n$ has multivariate lognormal distribution $\mathrm{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with density $f_{\mathrm{LN}}(\boldsymbol{w} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \exp\{-\frac{1}{2}(\log \boldsymbol{w} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\log \boldsymbol{w} - \boldsymbol{\mu})\} \prod_{i=1}^n \frac{1}{w_i} \mathrm{I}(w_i > 0)$.
- A random variable $W \in (0, \infty)$ has inverse gamma distribution $\mathrm{IG}(a, b)$ with density $f_{\mathrm{IG}}(w \mid a, b) \propto w^{-a-1} e^{-b/w} \mathrm{I}(w > 0)$.
- Random variable $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ represents the agency noise, with density $\xi_i \sim f_i(\xi_i \mid \lambda_i)$ and λ_i a known parameter. Density f_i will be taken to be the density of either $\mathrm{DG}(0, \lambda_i^2)$ or $\mathrm{N}(0, \lambda_i^2)$.

A.1 Gibbs Sampler for REG and IID

Let $\boldsymbol{X} \in \mathbb{R}^{n \times d}$ be an $n \times d$ design matrix corresponding to fixed effects $\boldsymbol{\beta}$. In vector notation, assume the REG model

$$\begin{aligned} &\boldsymbol{Z} = \boldsymbol{Y} + \boldsymbol{\xi}, \quad \xi_i \overset{\text{ind.}}{\sim} f_i(\xi_i \mid \lambda_i), \\ &\log(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}), \end{aligned}$$

with parameter model

$$\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I}), \quad \sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma).$$

The IID model is obtained as a special case of the REG model where $\boldsymbol{X} = \boldsymbol{1}_n$. It is straightforward to derive a Gibbs sampler for the REG model when $\boldsymbol{Y}$ is drawn as augmented data. However, because the distribution of $[\boldsymbol{Y} \mid \cdot]$ does not have a familiar conjugate form, we will make use of a direct sampler described in Appendix A.3.

The joint distribution of all random variables in the REG model is

$$\begin{aligned} f(\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) &= f(\boldsymbol{z} \mid \boldsymbol{y}) f(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}) f(\sigma^2) \\ &= \left[\prod_{i=1}^n f_i(z_i - y_i \mid \lambda_i) \right] f_{\mathrm{LN}}(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}) f_{\mathrm{N}}(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I}) f_{\mathrm{IG}}(\sigma^2 \mid a_\sigma, b_\sigma). \end{aligned}$$

1. To draw $[\sigma^2 \mid \cdot]$,

$$\begin{aligned} f(\sigma^2 \mid \cdot) &\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} (\log \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \underbrace{(\log \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}_{\boldsymbol{\eta}} \right\} \cdot (\sigma^2)^{-a_\sigma - 1} e^{-b_\sigma/\sigma^2} \\ &= (\sigma^2)^{-a_\sigma - n/2 - 1} \exp\left\{ -\frac{1}{\sigma^2} [b_\sigma + \boldsymbol{\eta}^\top \boldsymbol{\eta}/2] \right\}. \end{aligned}$$

   Therefore $[\sigma^2 \mid \cdot] \sim \mathrm{IG}(a_\sigma + n/2, b_\sigma + \boldsymbol{\eta}^\top \boldsymbol{\eta}/2)$.

2. To draw $[\boldsymbol{\beta} \mid \cdot]$,

$$f(\boldsymbol{\beta} \mid \cdot) \propto \exp\left\{ -\frac{1}{2\sigma^2} (\log \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\log \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right\} e^{-\boldsymbol{\beta}^\top \boldsymbol{\beta}/(2\sigma_\beta^2)}$$

   Therefore $[\boldsymbol{\beta} \mid \cdot] \sim \mathrm{N}(\boldsymbol{\vartheta}, \boldsymbol{\Omega}^{-1})$ with

$$\boldsymbol{\vartheta} = \boldsymbol{\Omega}^{-1} \boldsymbol{X}^\top (\log \boldsymbol{y}) \sigma^{-2}, \quad \boldsymbol{\Omega} = \sigma^{-2} \boldsymbol{X}^\top \boldsymbol{X} + \sigma_\beta^{-2} \boldsymbol{I}.$$

3. Finally, to draw $[\boldsymbol{Y} \mid \boldsymbol{z}, \boldsymbol{\theta}]$, notice that

$$f(\boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{\theta}) \propto \left[\prod_{i=1}^{n} f_i(z_i - y_i \mid \lambda_i) f_{\mathrm{LN}}(y_i \mid \boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2) \right]$$

so that Y_i may be drawn independently from $[\boldsymbol{Y} \mid \boldsymbol{z}, \boldsymbol{\theta}] = \prod_{i=1}^{n} [Y_i \mid z_i, \boldsymbol{\theta}]$. We may write

$$f(y_i \mid z_i, \boldsymbol{\theta}) \propto f_i(z_i - y_i \mid \lambda_i) f_{\mathrm{LN}}(y_i \mid \boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2). \tag{A.1}$$

However, rather than directly making use of (A.1), it is equivalent and conceptually helpful in the direct sampler to draw ξ_i' from

$$f(\xi_i \mid \cdot) \propto f_i(\xi_i \mid \lambda_i) f_{\mathrm{LN}}(z_i - \xi_i \mid \boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2),$$

as described in Appendix A.3, and transform to $y_i = z_i - \xi_i'$ to obtain a draw from $[Y_i \mid z_i, \boldsymbol{\theta}]$.

Remark A.1. When both z_i and ξ_i are integer-valued, density (A.1) will assign probability zero to the event $[y_i = 0]$, so that zeros are not drawn. In this case, it would be preferable to model the transformation $\log(\boldsymbol{Y} + \boldsymbol{1}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}$ rather than $\log(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}$.

A.2 Gibbs Sampler for CAR

Let us rewrite the CAR model as

$$\begin{aligned} &\boldsymbol{Z} = \boldsymbol{Y} + \boldsymbol{\xi}, \quad \xi_i \overset{\text{ind.}}{\sim} f_i(\xi_i \mid \lambda_i), \\ &\log(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} | \sigma^2 \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{R}^{-1}). \end{aligned}$$

with $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$ and $\boldsymbol{R} = \boldsymbol{D} - \rho \boldsymbol{A}$ with matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ as defined in Section 2.2, and assume the prior

$$\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I}), \quad \sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma).$$

If $\rho = 1$ is taken so that an Intrinsic CAR is assumed, a pseudo-inverse $\boldsymbol{R}^-$ may be used in place of $\boldsymbol{R}^{-1}$ when expressing the model. To derive a Gibbs sampler for this model, the joint distribution of all random variables is

$$\begin{aligned} f(\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2) &= f(\boldsymbol{z} \mid \boldsymbol{y}) f(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}) f(\sigma^2) \\ &= \left[\prod_{i=1}^{n} f_i(z_i - y_i \mid \lambda_i) \right] f_{\mathrm{LN}}(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{R}^{-1}) f_{\mathrm{N}}(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I}) f_{\mathrm{IG}}(\sigma^2 \mid a_\sigma, b_\sigma). \end{aligned}$$

The conditionals are obtained as follows.

1. To draw $[\sigma^2 \mid \cdot]$,

$$\begin{aligned} f(\sigma^2 \mid \cdot) &\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \underbrace{(\log \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{R} (\log \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}_{Q} \right\} \cdot (\sigma^2)^{-a_\sigma - 1} e^{-b_\sigma / \sigma^2} \\ &= (\sigma^2)^{-a_\sigma - n/2 - 1} \exp\left\{ -\frac{1}{\sigma^2} [b_\sigma + Q/2] \right\}. \end{aligned}$$

Therefore $[\sigma^2 \mid \cdot] \sim \mathrm{IG}(a_\sigma + n/2, b_\sigma + Q/2)$.

2. To draw $[\boldsymbol{\beta} \mid \cdot]$,

$$\begin{aligned} f(\boldsymbol{\beta} \mid \cdot) &\propto \exp\left\{-\frac{1}{2\sigma^2}(\log \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{R}(\log \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\} e^{-\boldsymbol{\beta}^\top\boldsymbol{\beta}/(2\sigma_\beta^2)} \\ &\propto \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\beta}^\top \underbrace{\left(\sigma^{-2}\boldsymbol{X}^\top\boldsymbol{R}\boldsymbol{X} + \sigma_\beta^{-2}\boldsymbol{I}\right)}_{\boldsymbol{\Omega}} \boldsymbol{\beta} - 2\sigma^{-2}(\log \boldsymbol{y})^\top \boldsymbol{R}\boldsymbol{X}\boldsymbol{\beta}\Big]\Big\} \\ &\propto \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\beta}^\top\boldsymbol{\Omega}\boldsymbol{\beta} - 2\underbrace{\sigma^{-2}(\log \boldsymbol{y})^\top \boldsymbol{R}\boldsymbol{X}\boldsymbol{\Omega}^{-1}}_{\boldsymbol{\vartheta}^\top}\boldsymbol{\Omega}\boldsymbol{\beta}\Big]\Big\}. \end{aligned}$$

Therefore $[\boldsymbol{\beta} \mid \cdot] \sim \mathrm{N}(\boldsymbol{\vartheta}, \boldsymbol{\Omega}^{-1})$ with

$$\boldsymbol{\vartheta} = \boldsymbol{\Omega}^{-1}\boldsymbol{X}^\top\boldsymbol{R}\,(\log \boldsymbol{y})\sigma^{-2}, \quad \boldsymbol{\Omega} = \sigma^{-2}\boldsymbol{X}^\top\boldsymbol{R}\boldsymbol{X} + \sigma_\beta^{-2}\boldsymbol{I}.$$

3. Finally, recalling that $\boldsymbol{\theta} = (\mu, \sigma^2)$, to draw $[\boldsymbol{Y} \mid \boldsymbol{z}, \boldsymbol{\theta}]$, notice that

$$f(\boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{\theta}) \propto \left[\prod_{i=1}^n f_i(z_i - y_i \mid \lambda_i)\right] f_{\mathrm{LN}}(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{R}) \tag{A.2}$$

may not factor into independent terms, however, we may draw sequentially from $[Y_i \mid \boldsymbol{y}_{-i}, \boldsymbol{z}, \boldsymbol{\theta}]$, for $i = 1, \ldots, n$, where $\boldsymbol{y}_{-i} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$. From (2.1) and (A.2),

$$\begin{aligned} f(y_i \mid \boldsymbol{y}_{-i}, \boldsymbol{z}, \boldsymbol{\theta}) &\propto f_i(z_i - y_i \mid \lambda_i) f_{\mathrm{LN}}(y_i \mid \mu_i, \sigma_i^2), \\ \mu_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \frac{\rho}{a_{i+}}\sum_{\ell=1}^n a_{i\ell}(\log y_\ell - \boldsymbol{x}_\ell^\top\boldsymbol{\beta}), &\quad \sigma_i^2 = \frac{\sigma^2}{a_{i+}}. \end{aligned}$$

Furthermore, rewriting as a distribution on ξ_i,

$$f(\xi_i \mid \boldsymbol{y}_{-i}, \boldsymbol{z}, \boldsymbol{\theta}) \propto f_i(\xi_i \mid \lambda_i) f_{\mathrm{LN}}(z_i - \xi_i \mid \mu_i, \sigma_i^2) \tag{A.3}$$

so that we may draw $\xi_i' \sim [\xi_i \mid \boldsymbol{y}_{-i}, \boldsymbol{z}, \boldsymbol{\theta}]$ using the direct sampler in Appendix A.3 and transform to $y_i = z_i - \xi_i'$ to obtain a draw from $[Y_i \mid \boldsymbol{y}_{-i}, \boldsymbol{z}, \boldsymbol{\theta}]$.

Remark A.2. With an ICAR model and an intercept assumed so that $\boldsymbol{X} = (\mathbf{1}\ \boldsymbol{W})$, this sampler cannot produce useful draws of the first element β_1 of $\boldsymbol{\beta}$. To see this, note that in Step 2, we have

$$\boldsymbol{R}\boldsymbol{X} = \begin{pmatrix}\mathbf{0} & (\boldsymbol{A} - \boldsymbol{D})\boldsymbol{W}\end{pmatrix} \quad \text{and} \quad \boldsymbol{X}^\top\boldsymbol{R}\boldsymbol{X} = \begin{pmatrix}0 & \mathbf{0}^\top \\ \mathbf{0} & \boldsymbol{W}^\top(\boldsymbol{A} - \boldsymbol{D})\boldsymbol{W}\end{pmatrix}$$

using the identities $(\boldsymbol{A} - \boldsymbol{D})\mathbf{1} = \mathbf{0}$ and $\mathbf{1}^\top(\boldsymbol{A} - \boldsymbol{D}) = \mathbf{0}^\top$. Furthermore,

$$\boldsymbol{\Omega}^{-1} = \begin{pmatrix}\sigma_\beta^2 & \mathbf{0}^\top \\ \mathbf{0} & \left[\sigma^{-2}\boldsymbol{W}^\top(\boldsymbol{A} - \boldsymbol{D})\boldsymbol{W} + \sigma_\beta^{-2}\boldsymbol{I}\right]^{-1}\end{pmatrix}.$$

Therefore, β_1 is drawn from a normal distribution independently of the remaining coordinates of $\boldsymbol{\beta}$ with variance $\boldsymbol{e}_1^\top\boldsymbol{\Omega}^{-1}\boldsymbol{e}_1 = \sigma_\beta^2$ and mean

$$\boldsymbol{e}_1^\top\boldsymbol{\vartheta} = \sigma^{-2}\boldsymbol{e}_1^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}^\top\boldsymbol{R}\,(\log \boldsymbol{y}) = \sigma^{-2}\sigma_\beta^2\boldsymbol{e}_1^\top\boldsymbol{X}^\top\boldsymbol{R}\,(\log \boldsymbol{y}) = 0,$$

which is a function of only the prior variance σ_β^2.

A.3 Direct Sampler

In Sections A.1 and A.2, we encountered conditional distributions for ξ_i of the form

$$f(\xi \mid \boldsymbol{\zeta}) = w(\xi \mid \boldsymbol{\zeta}) g(\xi \mid \lambda)/\psi, \quad \xi \in \Omega, \quad \psi = \int_\Omega w(\xi \mid \boldsymbol{\zeta}) g(\xi \mid \lambda) d\nu(\xi)$$

where $\nu(\cdot)$ is either the counting measure or Lesbegue measure if ξ_i is discrete or continuous, respectively. The vector $\boldsymbol{\zeta}$ contains all random variables which have been conditioned on; we will consider it fixed for the remainder of this section and omit it from the notation. Here, g is a density function which will be referred to as the base distribution, and w is a nonnegative function that will be referred as the weight function. In particular,

$$w(\xi \mid \boldsymbol{\zeta}) = f_{\text{LN}}(z - \xi \mid \vartheta, \tau^2) \tag{A.4}$$

for some mean and variance parameters ϑ and τ^2. Three base distributions were encountered, depending on the agency noise assumed in the model:

1. Laplace noise $g(\xi \mid \lambda) = f_{\text{Lap}}(\xi \mid \lambda)$,
2. Gaussian noise $g(\xi \mid \lambda) = f_{\text{N}}(\xi \mid 0, \lambda^2)$, and
3. Discrete Gaussian noise $g(\xi \mid \lambda) = f_{\text{DG}}(\xi \mid 0, \lambda^2)$.

These three weighted distributions may be drawn using the direct sampling scheme described in Raim (2021), which is based on the framework developed by Walker et al. (2011). The idea in direct sampling is to augment a random variable U where it is easier to draw from the joint distribution $[\xi, U]$ than the marginal of ξ. This is accomplished without explicitly computing the normalizing constant ψ. Define $c = \sup_{\xi \in \Omega} w(\xi)$ and $A_u = \{\xi \in \Omega : w(\xi) > uc\}$. To augment U, assume $[U \mid \xi] \sim \text{Uniform}(0, w(\xi)/c)$ so that

$$f(u \mid \xi) = \frac{c}{w(\xi)} \text{I}(0 < u < w(\xi)/c) = \frac{c}{w(\xi)} \text{I}(\xi \in A_u).$$

The joint density of $[\xi, U]$ is then

$$f(x, u) = \frac{c}{\psi} g(\xi) \text{I}(\xi \in A_u),$$

so that the marginal density of U is

$$p(u) = \frac{c}{\psi} \text{P}(A_u), \quad u \in [0, 1], \quad \text{P}(A_u) = \int \text{I}(\xi \in A_u) g(\xi) d\nu(\xi),$$

and the distribution of $[\xi \mid u]$ is

$$f(\xi \mid u) = \frac{g(\xi)}{\text{P}(A_u)} \text{I}(\xi \in A_u).$$

Therefore, we may draw u from marginal distribution $p(u)$ and ξ' from $[\xi \mid u]$, and finally discard u.

The direct sampling scheme described in Raim (2021) assumes univariate target density f with weight function w with A_u an interval for every $u \in [0, 1]$, and is most practical when endpoints of A_u and the CDF and quantile function of base distribution g are readily computed. We will review these steps to obtain samplers used the present paper. A step function $h(u)$ to approximate the marginal $p(u)$; in particular, an unnormalized version of the step function h^* is built from N knot points $u_0, \ldots, u_N$ so that $h(u) = h^*(u)/\int_0^1 h^*(t)dt$. We use $N = 100$ in all results in this paper. To obtain exact draws from $p(u)$, rather than from a numerical approximation to $p(u)$, rejection sampling is used with h as the proposal density as described in Algorithm 2. The expected number of rejections will be small when h^* is an accurate approximation to $\text{P}(A_u)$. Details on construction of the step function, knot selection, and rejection sampling are described in Raim (2021).

For the weight function (A.4), the mode is $\xi^* = z - \exp\{\mu - \sigma^2\}$ so that $c = \exp\{-(\mu - \sigma^2/2)\}$ is the associated maximum of the weight function. The set A_u is an interval whose endpoints are the roots of the equation $w(\xi) - cu = 0$, which are given by

$$\{\xi_1(u), \xi_2(u)\} = z - \exp\left\{(\mu - \sigma^2) \pm \left[\sigma^4 - 2\mu\sigma^2 + 2\sigma^2 \log(cu)\right]^{1/2}\right\}. \tag{A.5}$$

Denoting G as the cumulative distribution function associated with g, and $G(\xi-) = \lim_{t \uparrow \xi} G(t)$, we have

$$\begin{aligned} \mathrm{P}(A_u) &= \mathrm{P}(w(\xi) > uc) \\ &= \mathrm{P}(\xi_1(u) < \xi < \xi_2(u)) \\ &= G(\xi_2(u)-) - G(\xi_1(u)). \end{aligned} \tag{A.6}$$

Once u is drawn from $p(u)$, the quantile function $G^-(\xi)$ may be used to draw from $f(\xi \mid u)$ via the inverse CDF method because

$$F^-(\varphi \mid u) \equiv G^-((t-s)\varphi + s)$$

for quantile $\varphi \in (0,1)$ with $s = G(\xi_1(u))$ and $t = G(\xi_2(u)-)$. Algorithm 1 describes the computation of $\mathrm{P}(A_u)$ via (A.5) and (A.6), and Algorithm 2 summarizes the overall direct sampler.[4] To complete specification of the sampler, we now describe computation of G and G^- for each of the three base distributions. The Laplace distribution has closed form expressions

$$\begin{aligned} G_{\mathrm{Lap}}(\xi \mid \lambda) &= \frac{1}{2} + \frac{1}{2}\operatorname{sgn}(\xi)[1 - e^{-|\xi|/\lambda}], \\ G^-_{\mathrm{Lap}}(\varphi \mid \lambda) &= -\lambda \operatorname{sgn}(\varphi - 1/2)\log(1 - 2|\varphi - 1/2|). \end{aligned}$$

Functions for the Gaussian distribution are widely available in software libraries; e.g., $G_{\mathrm{N}}(\xi \mid \lambda^2)$ and $G^-_{\mathrm{N}}(\varphi \mid \lambda^2)$ can be invoked using `pnorm(xi, 0, lambda)` and `qnorm(phi, 0, lambda)` in R, respectively. The $\mathrm{DG}(0, \lambda^2)$ distribution is not implemented in standard statistical software; however, Canonne et al. (2020) provide a number of results to facilitate its use, including a close relationship to the $\mathrm{N}(0, \lambda^2)$ distribution. We will make use of one particular result, which will now be stated for reference.

Lemma A.3 (Proposition 25 of Canonne et al. (2020))**.** *Suppose $X \sim DG(0, \lambda^2)$ and $Y \sim N(0, \lambda^2)$. For any integer $m \geq 1$ and all $\lambda > 0$, $\mathrm{P}(X \geq m) \leq \mathrm{P}(Y \geq m - 1)$.*

To use the Discrete Gaussian distribution with the direct sampler, we would like to make use of its CDF and quantile function; however, these appear not to have closed form expressions. Our approach will be to work with a truncated version of the distribution, which is described in the following results. Denote $\Phi(x \mid \mu, \sigma^2)$ and $\Phi^{-1}(\varphi \mid \mu, \sigma^2)$ as the CDF and quantile function of $N(\mu, \sigma^2)$, respectively.

Lemma A.4. *For any $\tau \in (0, 1]$, let $x_\tau = \lceil \Phi^{-1}(1 - \tau/2 \mid 0, \lambda^2) \rceil$ and $X \sim DG(0, \lambda^2)$. Then $\mathrm{P}(|X| > x_\tau) \leq \tau$.*

Proof. Let $Y \sim \mathrm{N}(0, \lambda^2)$ and $U \sim \mathrm{Uniform}(0, 1)$,

$$\begin{aligned} \mathrm{P}(|X| > x_\tau) &= 2\,\mathrm{P}(X \geq x_\tau + 1) && \text{(A.7)} \\ &\leq 2\,\mathrm{P}(Y \geq x_\tau) && \text{(A.8)} \\ &\leq 2\,\mathrm{P}(Y \geq \Phi^{-1}(1 - \tau/2 \mid 0, \lambda^2)) \\ &= 2\,\mathrm{P}(U \geq 1 - \tau/2) \\ &= 2(1 - (1 - \tau/2)) \\ &= \tau. \end{aligned}$$

Expression (A.7) follows by symmetry of X about 0. Expression (A.8) follows from Lemma A.3 because x_τ is a nonnegative integer when $\tau \in (0, 1]$. □

[4] Remark 2.2 of Raim (2021) gives a more complicated expression for the acceptance ratio, but it can simply be taken as $\mathrm{P}(A_u)/h^*(u)$.

Now, with x_τ selected as in Lemma A.4, $S_\tau = \{-x_\tau, \ldots, x_\tau\}$ is ensured to contain at least probability τ from the $\mathrm{DG}(0, \lambda^2)$ distribution. Therefore, we consider instead working with $\mathrm{DG}(0, \lambda^2)$ truncated to the set S_τ, which we denote $\mathrm{DG}_\tau(0, \lambda^2)$. The following result shows that $\mathrm{DG}(0, \lambda^2)$ and $\mathrm{DG}_\tau(0, \lambda^2)$ are also close in total variation distance d_{TV}.

Proposition A.5. *Let p be a density on a countable set Ω, $X \sim p$, and $q(x) = \mathrm{I}(x \in S) \cdot p(x) / \mathrm{P}(X \in S)$ be q truncated to a set $S \subseteq \Omega$. Then $d_{TV}(p, q) = \mathrm{P}(X \notin S)$.*

Proof. Because Ω is countable, we may express the total variation distance between p and q as

$$\begin{aligned}
d_{\mathrm{TV}}(p, q) &= \frac{1}{2} \sum_{x \in \Omega} |p(x) - q(x)| \\
&= \frac{1}{2} \sum_{x \in S} |p(x) - q(x)| + \frac{1}{2} \sum_{x \in \Omega \setminus S} |p(x) - q(x)| \\
&= \frac{1}{2} \sum_{x \in S} \left| p(x) - \frac{p(x)}{\mathrm{P}(X \in S)} \right| + \frac{1}{2} \sum_{x \in \Omega \setminus S} p(x) \\
&= \frac{1}{2} \sum_{x \in S} p(x) \left| 1 - \frac{1}{\mathrm{P}(X \in S)} \right| + \frac{1}{2} \mathrm{P}(X \notin S) \\
&= \frac{1}{2} \sum_{x \in S} p(x) \frac{\mathrm{P}(X \notin S)}{\mathrm{P}(X \in S)} + \frac{1}{2} \mathrm{P}(X \notin S) \\
&= \mathrm{P}(X \notin S).
\end{aligned}$$

□

Therefore, the total variation distance between $\mathrm{DG}(0, \lambda^2)$ and $\mathrm{DG}_\tau(0, \lambda^2)$ is $\mathrm{P}(|X| > x_\tau)$, which is bounded above by the τ given in Lemma A.4 to select x_τ. To facilitate the repeated computations of the CDF of $\mathrm{DG}_\tau(0, \lambda^2)$ used in Algorithms 1 and 2, values

$$G_{\mathrm{DG}_\tau}(\xi \mid \lambda) = \frac{\sum_{\omega \in S_\tau} \mathrm{I}(\omega \leq \xi) \exp\{-\omega^2/(2\lambda^2)\}}{\sum_{s \in S_\tau} \exp\{-s^2/(2\lambda^2)\}}$$

may be precomputed for a given λ and each $\xi \in S_\tau$. The quantile function $G^-_{\mathrm{DG}_\tau}$ may be implemented using a bisection search on the sorted CDF values.

Algorithm 1 Unnormalized $p(u)$ density.

Input: $u \in [0, 1]$ argument to evaluate.
Input: $\boldsymbol{\zeta} = (z, \mu, \sigma^2)$ parameters for w.

1: **function** $\mathrm{P}(A_u)$
2: Let $\{\xi_1(u), \xi_2(u)\} = z - \exp\left\{(\mu - \sigma^2) \pm \left[\sigma^4 - 2\mu\sigma^2 + 2\sigma^2 \log(cu)\right]^{1/2}\right\}$.
3: **return** $G(\xi_2(u)-) - G(\xi_1(u))$

B Simulation Results

Algorithm 2 Direct sampler with rejection step.

Input: G is CDF of g.
Input: G^- is quantile function of g.
Input: $c = \sup_{\xi \in \Omega} w(\xi)$.
Input: $\boldsymbol{\zeta} = (z, \mu, \sigma^2)$ parameters for w.
Input: λ parameter for g.
Input: N determines initial number of knot points for step function approximation.

function DIRECTSAMPLER
  Define the function $\mathrm{P}(A_u)$ according to Algorithm 1.
  Construct the step function h^* using $N+1$ knot points and $\mathrm{P}(A_u)$.
  Let h be the density from normalizing h^*.

  do ▷ Draw from $p(u)$ using rejection sampling.
    Draw u from step density h
    Draw v from Uniform$(0, 1)$
    Let Accept $\leftarrow \mathrm{I}[v \leq \mathrm{P}(A_u)/h^*(u)]$.
    Update h^* with u as additional knot if adaptive rejection is desired.
  while Accept $= 0$

  Assign $s = G(\xi_1(u))$ and $t = G(\xi_2(u)-)$ ▷ Draw from $f(x \mid u)$ using its quantile function.
  Draw $\omega \sim$ Uniform$(0, 1)$
  return $G^-((t-s)\omega + s)$

Table 1: Results for Laplace IID simulation: summary of predictions.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Observed	RMSE	140.81	14.16	2.82	141.07	14.15	2.84	141.41	14.07	2.81
	MAE	99.75	10.03	2.00	100.13	10.03	2.01	100.22	9.99	1.99
	MRAE	59.91	6.14	1.20	36.38	3.68	0.75	23.01	2.15	0.44
	MAX	764.34	75.58	15.11	736.72	73.74	15.06	743.83	73.77	14.38
	COV	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	LEN	599.15	59.91	11.98	599.15	59.91	11.98	599.15	59.91	11.98
Laplace	RMSE	7.75	5.05	2.04	40.61	9.77	2.50	77.01	11.66	2.64
	MAE	4.41	3.14	1.46	19.42	6.71	1.77	48.19	8.20	1.88
	MRAE	1.60	1.74	0.82	5.84	2.21	0.63	9.30	1.61	0.40
	MAX	97.19	47.69	12.90	559.13	61.41	13.87	562.19	68.41	14.12
	COV	0.42	0.94	0.95	0.87	0.94	0.95	0.92	0.95	0.95
	LEN	10.71	14.83	7.27	86.92	32.20	9.89	214.79	44.12	10.91
Gauss	RMSE	30.21	5.96	2.06	48.63	9.87	2.50	79.19	11.70	2.64
	MAE	11.99	3.38	1.45	23.93	6.67	1.77	48.05	8.18	1.88
	MRAE	7.89	1.84	0.81	8.38	2.18	0.61	9.49	1.58	0.39
	MAX	523.57	59.86	13.38	580.25	63.64	13.96	614.23	69.06	14.14
	COV	0.86	0.96	0.93	0.94	0.93	0.92	0.92	0.92	0.92
	LEN	72.39	16.55	6.63	122.02	29.48	8.36	211.23	37.78	9.06

Table 2: Results for Laplace IID simulation: summary of parameter estimates.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Laplace	μ	-35.08	1.00	1.00	2.26	2.04	2.01	3.23	3.03	3.01
	sd(μ)	64.54	0.12	0.05	0.26	0.07	0.05	0.14	0.06	0.06
	σ^2	1.07	0.99	1.01	1.26	1.88	1.94	2.40	2.77	2.86
	sd(σ^2)	0.18	0.15	0.08	0.26	0.15	0.12	0.23	0.15	0.13
Gauss	μ	-0.21	0.86	0.99	2.30	2.02	2.01	3.23	3.03	3.00
	sd(μ)	5.72	0.13	0.05	0.16	0.08	0.05	0.14	0.06	0.06
	σ^2	1.92	1.43	1.07	2.02	1.97	1.96	2.55	2.82	2.88
	sd(σ^2)	0.42	0.16	0.08	0.20	0.16	0.12	0.23	0.16	0.13

Table 3: Results for Laplace REG simulation: summary of predictions.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Observed	RMSE	142.21	14.04	2.83	141.69	14.14	2.83	141.78	14.09	2.84
	MAE	100.77	9.95	2.00	100.00	10.00	2.00	100.26	9.96	2.02
	MRAE	69.83	6.85	1.38	122.44	13.21	2.36	469.15	44.48	9.08
	MAX	746.00	74.27	14.71	773.21	73.52	14.72	765.87	75.50	14.92
	COV	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	LEN	599.15	59.91	11.98	599.15	59.91	11.98	599.15	59.91	11.98
Laplace	RMSE	8.89	5.44	2.03	54.38	9.70	2.40	85.09	11.15	2.60
	MAE	4.56	3.26	1.43	26.25	6.24	1.66	47.49	7.27	1.82
	MRAE	1.60	1.64	0.84	5.54	2.72	1.10	10.12	4.21	4.11
	MAX	112.44	48.00	12.32	511.17	62.82	13.65	654.78	69.80	14.22
	COV	0.67	0.94	0.95	0.92	0.94	0.95	0.94	0.94	0.79
	LEN	14.89	15.25	7.20	124.21	31.88	9.16	243.75	39.26	8.31
Gauss	RMSE	33.22	6.19	2.06	59.93	9.83	2.41	87.26	11.19	2.55
	MAE	12.79	3.48	1.43	29.03	6.26	1.66	48.37	7.25	1.76
	MRAE	9.21	1.82	0.84	9.53	3.05	1.12	14.11	4.38	2.66
	MAX	544.62	58.89	12.75	604.26	64.75	13.76	685.77	70.74	14.19
	COV	0.96	0.96	0.93	0.93	0.93	0.92	0.92	0.92	0.86
	LEN	79.20	16.83	6.55	137.10	28.71	7.82	223.27	34.03	7.71

Table 4: Results for Laplace REG simulation: summary of parameter estimates.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Laplace	β_0	-1.80	0.43	0.49	0.60	0.52	0.52	0.71	0.56	0.90
	$\text{sd}(\beta_0)$	2.54	0.19	0.07	0.41	0.16	0.09	0.28	0.21	0.34
	β_1	0.19	0.54	0.51	1.51	1.50	1.49	2.43	2.46	2.26
	$\text{sd}(\beta_1)$	1.12	0.09	0.04	0.17	0.08	0.06	0.14	0.13	0.21
	σ^2	1.04	1.00	1.01	1.66	1.91	1.95	2.64	2.79	2.67
	$\text{sd}(\sigma^2)$	0.15	0.28	0.16	0.65	0.56	0.42	1.12	0.85	1.25
Gauss	β_0	1.26	0.41	0.48	0.99	0.56	0.52	0.87	0.57	0.68
	$\text{sd}(\beta_0)$	0.83	0.19	0.07	0.45	0.16	0.09	0.28	0.17	0.29
	β_1	0.19	0.47	0.51	1.31	1.47	1.49	2.35	2.46	2.39
	$\text{sd}(\beta_1)$	0.39	0.09	0.04	0.20	0.08	0.06	0.14	0.11	0.18
	σ^2	2.04	1.41	1.07	1.98	1.98	1.98	2.75	2.84	2.79
	$\text{sd}(\sigma^2)$	1.31	0.45	0.18	0.78	0.59	0.43	1.17	0.89	0.99

Table 5: Results for Laplace CAR simulation: summary of predictions.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Observed	RMSE	140.41	14.20	2.81	141.81	14.35	2.81	139.57	14.12	2.81
	MAE	99.90	10.08	1.99	100.82	10.16	2.00	99.31	10.05	2.00
	MRAE	44.85	4.63	0.94	47.46	4.79	0.89	85.59	9.00	1.65
	MAX	506.83	53.82	10.51	526.40	54.17	10.14	525.66	52.78	10.14
	COV	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	LEN	599.15	59.91	11.98	599.15	59.91	11.98	599.15	59.91	11.98
Laplace	RMSE	5.29	2.95	1.37	37.37	7.97	2.18	73.82	10.66	2.45
	MAE	3.50	2.02	0.97	17.45	4.87	1.51	39.25	6.76	1.70
	MRAE	1.37	0.64	0.36	1.82	0.54	0.30	1.05	0.55	0.35
	MAX	23.98	13.56	5.58	226.08	35.09	8.38	350.15	45.15	9.19
	COV	0.88	0.91	0.94	0.91	0.92	0.94	0.90	0.91	0.90
	LEN	17.51	8.22	4.75	70.40	23.44	8.11	193.27	34.75	8.96
Gauss	RMSE	6.84	3.63	1.44	38.44	8.14	2.19	72.66	10.64	2.41
	MAE	4.06	2.41	1.01	19.27	4.98	1.51	39.09	6.74	1.65
	MRAE	1.58	0.80	0.38	2.35	0.66	0.31	1.22	0.56	0.31
	MAX	38.00	17.96	6.09	224.92	36.74	8.54	348.04	45.61	9.12
	COV	0.89	0.86	0.94	0.89	0.90	0.92	0.88	0.90	0.90
	LEN	20.63	8.81	4.77	75.97	22.04	7.11	173.29	30.44	7.75

Table 6: Results for Laplace CAR simulation: summary of parameter estimates.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Laplace	β_0	-3.30	-0.79	0.45	-0.56	0.46	0.53	0.66	0.57	0.59
	sd(β_0)	1.04	1.57	0.25	1.70	0.41	0.31	0.63	0.45	0.47
	β_1	-1.34	0.36	0.51	1.48	1.51	1.50	2.44	2.46	2.40
	sd(β_1)	0.70	0.58	0.10	0.84	0.16	0.09	0.24	0.14	0.11
	σ^2	1.00	1.00	1.00	1.04	1.38	1.60	1.66	2.02	2.21
	sd(σ^2)	0.02	0.05	0.15	0.08	0.23	0.19	0.25	0.26	0.27
Gauss	β_0	-3.22	-1.09	0.46	-0.36	0.56	0.55	0.76	0.66	0.51
	sd(β_0)	1.20	1.97	0.28	2.03	0.46	0.30	0.63	0.44	0.38
	β_1	-1.30	0.29	0.50	1.47	1.47	1.49	2.41	2.45	2.45
	sd(β_1)	0.77	0.80	0.11	0.89	0.19	0.09	0.25	0.15	0.10
	σ^2	1.01	1.11	1.11	1.12	1.50	1.67	1.72	2.06	2.22
	sd(σ^2)	0.11	0.21	0.20	0.17	0.27	0.21	0.29	0.26	0.24

Table 7: Results for DG IID simulation: summary of predictions.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Observed	RMSE	100.07	10.02	2.00	100.07	10.02	2.00	100.07	10.02	2.00
	MAE	79.82	7.99	1.57	79.82	7.99	1.57	79.82	7.99	1.57
	MRAE	48.75	4.80	0.95	29.68	2.88	0.57	18.07	1.72	0.34
	MAX	347.97	34.14	6.84	347.97	34.14	6.84	347.97	34.14	6.84
	COV	0.95	0.95	0.93	0.95	0.95	0.93	0.95	0.95	0.92
	LEN	392.00	40.00	8.00	392.00	40.00	8.00	392.00	40.00	8.00
DG	RMSE	8.83	4.20	1.55	31.01	7.25	1.82	56.39	8.57	1.92
	MAE	6.82	2.82	1.19	17.84	5.36	1.44	37.07	6.59	1.51
	MRAE	4.89	1.51	0.66	5.67	1.67	0.48	6.48	1.22	0.31
	MAX	88.58	27.82	6.36	285.26	31.46	6.69	317.77	32.90	6.77
	COV	1.00	0.98	0.98	0.97	0.97	0.97	0.97	0.96	0.97
	LEN	51.04	14.02	5.51	95.75	25.14	6.67	176.70	31.34	7.34
Gauss	RMSE	7.06	4.50	1.59	35.16	7.43	1.84	59.15	8.65	14.58
	MAE	4.08	2.79	1.23	16.08	5.52	1.46	37.80	6.72	3.48
	MRAE	0.86	1.33	0.72	2.73	1.80	0.54	6.28	1.37	0.36
	MAX	70.57	32.64	6.64	347.85	32.99	6.74	344.98	33.44	350.86
	COV	0.39	0.95	0.97	0.89	0.97	0.98	0.96	0.97	0.96
	LEN	3.32	13.31	6.57	59.98	29.19	8.40	193.68	37.84	9.80

Table 8: Results for DG IID simulation: summary of parameter estimates.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
DG	μ	1.45	0.93	0.99	2.03	1.99	2.00	2.92	3.00	3.00
	sd(μ)	0.21	0.11	0.04	0.19	0.08	0.05	0.15	0.07	0.05
	σ^2	1.60	1.15	1.04	2.03	2.07	2.01	3.20	3.05	2.99
	sd(σ^2)	0.15	0.13	0.08	0.27	0.19	0.12	0.33	0.20	0.17
Gauss	μ	-5.04	0.79	0.99	0.24	1.98	2.00	2.77	3.01	2.98
	sd(μ)	2.44	0.16	0.05	1.80	0.09	0.05	0.22	0.07	0.10
	σ^2	5.11	1.05	0.97	3.68	2.00	1.96	3.29	2.99	3.05
	sd(σ^2)	1.71	0.19	0.08	2.21	0.20	0.12	0.47	0.21	0.39

Table 9: Results for DG REG simulation: summary of predictions.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Observed	RMSE	99.84	9.99	1.99	99.84	9.99	1.99	99.84	9.99	1.99
	MAE	79.58	7.98	1.56	79.58	7.98	1.56	79.58	7.98	1.56
	MRAE	55.82	5.55	1.08	96.42	9.69	1.82	403.25	41.19	6.94
	MAX	341.50	34.04	6.81	341.50	34.04	6.81	341.50	34.04	6.81
	COV	0.95	0.95	0.93	0.95	0.95	0.93	0.95	0.95	0.93
	LEN	392.00	40.00	8.00	392.00	40.00	8.00	392.00	40.00	8.00
DG	RMSE	9.56	4.36	1.53	40.44	7.16	1.73	60.88	8.06	1.78
	MAE	6.91	2.87	1.16	21.76	5.03	1.32	37.35	5.80	1.34
	MRAE	5.15	1.51	0.67	4.95	1.97	0.68	7.05	2.28	0.75
	MAX	100.32	27.87	6.44	286.64	32.05	6.70	317.51	32.59	6.81
	COV	1.00	0.98	0.98	0.97	0.97	0.98	0.97	0.97	0.98
	LEN	51.23	14.12	5.34	110.22	24.03	6.28	182.53	27.87	6.55
Gauss	RMSE	8.74	4.37	1.53	40.53	7.16	1.74	60.92	8.13	1.83
	MAE	4.46	2.84	1.16	21.49	5.04	1.32	37.38	5.89	1.40
	MRAE	1.27	1.45	0.68	4.82	2.21	0.85	8.63	4.79	2.42
	MAX	104.61	27.98	6.44	287.46	32.05	6.70	318.19	32.77	6.93
	COV	0.63	0.96	0.95	0.95	0.95	0.95	0.95	0.93	0.86
	LEN	12.72	14.24	5.58	107.69	24.17	6.42	182.31	27.74	6.14

Table 10: Results for DG REG simulation: summary of parameter estimates.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
DG	β_0	1.21	0.47	0.48	0.54	0.50	0.49	0.50	0.50	0.49
	$\text{sd}(\beta_0)$	0.27	0.13	0.06	0.14	0.10	0.07	0.11	0.09	0.08
	β_1	0.19	0.49	0.51	1.47	1.50	1.50	2.50	2.50	2.51
	$\text{sd}(\beta_1)$	0.15	0.07	0.04	0.07	0.05	0.05	0.05	0.06	0.05
	σ^2	1.62	1.10	1.04	2.09	2.02	2.02	3.03	3.00	3.00
	$\text{sd}(\sigma^2)$	0.15	0.10	0.06	0.22	0.14	0.12	0.20	0.17	0.16
Gauss	β_0	-3.63	0.34	0.47	0.27	0.49	0.48	0.46	0.57	0.68
	$\text{sd}(\beta_0)$	2.45	0.21	0.07	0.49	0.16	0.09	0.27	0.27	0.32
	β_1	-0.06	0.52	0.51	1.57	1.50	1.51	2.52	2.46	2.39
	$\text{sd}(\beta_1)$	0.96	0.09	0.04	0.19	0.09	0.06	0.13	0.16	0.19
	σ^2	4.18	1.20	1.05	2.22	2.03	2.02	3.05	3.00	2.99
	$\text{sd}(\sigma^2)$	1.48	0.16	0.07	0.37	0.18	0.13	0.25	0.19	0.55

Table 11: Results for DG CAR simulation: summary of predictions.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
Observed	RMSE	99.48	9.92	1.99	99.48	9.92	1.99	99.48	9.92	1.99
	MAE	79.43	7.94	1.56	79.43	7.94	1.56	79.43	7.94	1.56
	MRAE	35.44	3.57	0.72	34.61	3.61	0.76	63.52	6.95	1.56
	MAX	270.89	26.78	5.44	270.89	26.78	5.44	270.89	26.78	5.44
	COV	0.95	0.95	0.93	0.95	0.95	0.93	0.95	0.95	0.93
	LEN	392.00	40.00	8.00	392.00	40.00	8.00	392.00	40.00	8.00
DG	RMSE	7.81	2.52	1.16	30.13	6.03	1.63	55.42	7.47	1.74
	MAE	6.12	1.70	0.86	15.10	4.00	1.20	31.09	5.20	1.29
	MRAE	3.21	0.60	0.30	1.62	0.46	0.22	0.77	0.38	0.18
	MAX	31.75	11.40	3.96	168.06	22.31	4.96	224.16	24.79	5.17
	COV	1.00	0.99	0.99	0.99	0.97	0.98	0.97	0.97	0.98
	LEN	46.28	10.69	4.37	85.84	20.26	5.90	155.70	25.69	6.37
Gauss	RMSE	5.06	2.78	1.17	31.13	6.06	1.63	55.76	7.50	1.75
	MAE	3.34	1.90	0.86	15.05	4.02	1.21	31.56	5.24	1.31
	MRAE	1.10	0.55	0.30	1.21	0.47	0.25	0.97	0.45	0.27
	MAX	24.08	12.04	3.98	172.99	22.43	4.96	224.47	24.86	5.21
	COV	0.85	0.94	0.97	0.97	0.96	0.95	0.96	0.95	0.94
	LEN	14.86	9.98	4.61	79.84	20.43	6.02	157.29	25.80	6.39

Table 12: Results for DG CAR simulation: summary of parameter estimates.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
DG	β_0	1.28	0.48	0.44	0.85	0.53	0.47	0.55	0.52	0.47
	sd(β_0)	0.49	0.31	0.20	0.37	0.29	0.27	0.33	0.32	0.33
	β_1	0.05	0.36	0.50	1.28	1.49	1.51	2.48	2.50	2.51
	sd(β_1)	0.27	0.20	0.08	0.22	0.10	0.08	0.11	0.10	0.10
	σ^2	5.03	3.17	1.66	3.80	2.46	2.20	3.50	3.26	3.10
	sd(σ^2)	1.32	0.74	0.31	1.08	0.58	0.39	0.76	0.66	0.52
Gauss	β_0	-3.85	-0.94	0.37	-0.75	0.39	0.46	0.34	0.44	0.46
	sd(β_0)	0.79	1.38	0.27	1.31	0.41	0.29	0.68	0.43	0.42
	β_1	-1.45	0.55	0.53	1.69	1.55	1.52	2.57	2.51	2.45
	sd(β_1)	0.56	0.55	0.09	0.63	0.15	0.09	0.26	0.15	0.11
	σ^2	7.85	4.69	1.77	4.74	2.56	2.23	3.60	3.29	3.10
	sd(σ^2)	1.85	1.62	0.35	1.52	0.61	0.40	0.84	0.70	0.54

Table 13: Laplace simulation results: mean of STAN run times per simulation (minutes) over 100 simulated datasets.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
IID	Lap	0.64	1.01	4.05	1.41	3.57	4.06	2.32	3.37	5.59
	Gauss	0.39	0.17	0.14	0.27	0.17	0.22	0.24	0.26	0.33
REG	Lap	0.76	1.77	2.84	2.95	4.72	6.33	7.19	32.40	49.62
	Gauss	0.30	0.19	0.18	0.28	0.33	0.44	0.52	1.94	3.67
CAR	Lap	1.14	1.80	1.48	4.75	19.86	83.33	87.25	139.65	146.71
	Gauss	0.82	1.14	0.59	1.61	2.61	7.16	7.89	16.93	17.82

Table 14: DG simulation results: mean of sampler run times per simulation (minutes) over 100 simulated datasets.

		Scenario S1			Scenario S2			Scenario S3		
		$\epsilon = 0.01$	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
IID	DG	14.19	12.09	11.61	13.94	11.95	13.53	13.72	15.11	11.75
	Gauss	0.53	0.25	0.23	0.40	0.28	0.38	0.36	1.41	30.52
REG	DG	18.01	16.19	18.39	17.59	15.92	16.63	17.36	16.58	16.05
	Gauss	0.37	0.16	0.19	0.24	0.29	0.48	0.58	2.23	3.28
CAR	DG	1.63	1.34	1.23	1.57	1.35	1.23	1.61	1.28	1.19
	Gauss	0.94	1.10	0.66	1.87	3.69	9.03	8.99	16.72	12.78

Table 15: Laplace empirical simulation results: mean of STAN run times per simulation (minutes) over 100 simulated datasets.

		Cherokee			Choctaw			Sioux		
		$\epsilon = 0.01$	0.10	0.50	0.01	0.10	0.50	0.01	0.10	0.50
IID	Lap	0.21	0.23	0.97	0.18	0.15	0.34	0.10	0.21	0.17
	Gauss	0.03	0.05	0.10	0.02	0.03	0.05	0.05	0.02	0.02
REG	Lap	0.15	0.12	0.43	0.13	0.08	0.18	0.06	0.04	0.05
	Gauss	0.14	0.06	0.07	0.12	0.05	0.05	0.07	0.03	0.02
ICAR	Gauss	13.69	63.72	535.93	...	...	...	...	...	...

Table 16: DG empirical simulation results: mean of sampler run times per simulation (minutes) over 100 simulated datasets.

		Scenario 1			Scenario 2			Scenario 3		
		$\epsilon = 0.01$	0.10	0.50	0.01	0.10	0.50	0.01	0.10	0.50
IID	DG	9.96	8.26	8.44	9.73	8.42	8.25	9.91	8.52	8.34
	Gauss	1.06	2.40	3.40	0.67	1.49	2.77	0.33	0.61	0.99
REG	DG	11.40	9.96	8.03	11.74	9.91	8.18	11.42	9.38	8.10
	Gauss	8.55	15.11	16.76	2.70	7.06	15.43	1.15	1.68	4.32
CAR	DG	13.08	11.45	9.74	15.74	13.57	9.61	16.30	15.93	9.54
	Gauss	10.74	10.68	10.56	10.39	10.60	10.58	10.43	10.60	10.63

Table 17: Empirical simulation results: summary of predictions. Due to convergence issues, results for the Laplace ICAR model are not shown. Furthermore, Gauss ICAR results are shown only for the Cherokee group because of long runtimes.

		Cherokee			Choctaw			Sioux		
		$\epsilon = 0.01$	0.10	0.50	0.01	0.10	0.50	0.01	0.10	0.50
Observed	RMSE	140.78	13.82	2.83	138.55	14.06	2.75	142.74	13.96	2.77
	MAE	101.02	9.92	2.00	98.99	10.05	1.96	101.58	9.94	1.97
	MRAE	3.09	0.30	0.06	1.62	0.16	0.03	23.67	2.33	0.47
	MAX	490.57	47.55	10.09	490.22	48.65	9.37	512.41	49.02	9.86
	COV	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	LEN	599.15	59.91	11.98	599.15	59.91	11.98	599.15	59.91	11.98
Lap IID	RMSE	120.18	13.45	2.81	119.23	13.52	2.73	34.14	9.61	2.44
	MAE	85.48	9.69	1.99	85.82	9.64	1.95	16.43	6.73	1.75
	MRAE	2.14	0.30	0.06	1.23	0.14	0.03	1.65	1.16	0.31
	MAX	434.28	46.97	10.04	420.74	47.08	9.35	198.34	37.89	8.97
	COV	0.95	0.96	0.97	0.95	0.95	0.97	0.84	0.98	0.98
	LEN	459.48	58.29	12.81	458.90	58.02	12.87	58.03	31.89	10.45
Gauss IID	RMSE	120.91	13.45	2.81	119.68	13.53	2.73	45.68	9.82	2.44
	MAE	85.86	9.68	1.99	85.92	9.64	1.95	20.13	6.77	1.73
	MRAE	2.10	0.29	0.05	1.22	0.14	0.03	2.64	1.16	0.31
	MAX	438.15	47.08	10.06	425.97	47.17	9.35	277.91	40.19	9.06
	COV	0.93	0.93	0.95	0.93	0.92	0.95	0.86	0.96	0.96
	LEN	393.83	48.24	10.71	391.41	48.08	10.77	74.42	29.57	8.91
Lap REG	RMSE	80.17	13.41	3.49	71.35	12.32	2.86	26.51	6.25	2.41
	MAE	51.63	9.80	2.35	45.48	9.03	2.08	11.95	4.04	1.78
	MRAE	0.28	0.22	0.12	0.21	0.10	0.03	0.84	0.47	0.30
	MAX	323.93	44.89	13.88	286.85	42.11	9.27	156.99	27.61	8.15
	COV	0.81	0.89	0.94	0.82	0.90	0.97	0.94	0.88	0.93
	LEN	224.36	48.71	12.94	189.58	46.63	12.55	52.09	17.01	8.50
Gauss REG	RMSE	81.22	13.11	2.89	72.96	12.29	2.74	47.66	6.73	2.33
	MAE	51.96	9.59	2.08	46.59	8.99	1.99	19.54	4.26	1.71
	MRAE	0.29	0.22	0.09	0.22	0.10	0.03	1.73	0.52	0.29
	MAX	332.93	45.12	10.00	294.56	42.57	9.18	301.35	30.37	8.35
	COV	0.81	0.87	0.93	0.84	0.88	0.95	0.48	0.94	0.98
	LEN	210.36	41.63	10.50	193.61	40.40	10.43	35.07	18.64	7.81
Gauss ICAR	RMSE	85.37	9.64	2.61	—	—	—	—	—	—
	MAE	54.03	6.36	1.93	—	—	—	—	—	—
	MRAE	0.35	0.42	0.21	—	—	—	—	—	—
	MAX	358.25	60.21	12.50	—	—	—	—	—	—
	COV	0.92	0.84	0.93	—	—	—	—	—	—
	LEN	257.70	25.16	11.22	—	—	—	—	—	—

Table 18: Empirical (DG) simulation results: summary of predictions.

		Cherokee			Choctaw			Sioux		
		$\epsilon = 0.01$	0.10	0.50	0.01	0.10	0.50	0.01	0.10	0.50
Observed	RMSE	99.94	9.99	1.99	99.94	9.99	1.99	99.94	9.99	1.99
	MAE	79.75	7.96	1.55	79.75	7.96	1.55	79.75	7.96	1.55
	MRAE	10.05	1.00	0.20	20.92	2.08	0.41	36.27	3.61	0.70
	MAX	337.49	33.25	6.63	337.49	33.25	6.63	337.49	33.25	6.63
	COV	0.95	0.95	0.92	0.95	0.95	0.92	0.95	0.95	0.92
	LEN	392.00	40.00	8.00	392.00	40.00	8.00	392.00	40.00	8.00
DG IID	RMSE	59.04	8.66	1.93	51.92	7.72	1.87	31.34	6.22	1.76
	MAE	40.98	6.69	1.54	36.08	5.91	1.52	20.04	4.86	1.46
	MRAE	4.25	0.67	0.17	6.44	1.01	0.26	6.54	1.53	0.41
	MAX	299.37	31.83	6.55	298.18	30.84	6.46	288.98	29.48	6.20
	COV	0.93	0.93	0.94	0.84	0.83	0.84	0.69	0.69	0.69
	LEN	197.44	31.64	7.30	168.47	26.67	6.41	91.72	19.60	5.15
Gauss IID	RMSE	58.86	8.66	1.92	50.06	7.60	1.83	29.28	5.84	1.62
	MAE	38.01	6.62	1.52	29.49	5.54	1.44	12.92	3.91	1.24
	MRAE	3.11	0.60	0.15	3.92	0.74	0.22	2.86	0.87	0.33
	MAX	306.48	32.10	6.57	307.47	31.40	6.51	281.42	30.35	6.33
	COV	0.93	0.92	0.91	0.83	0.82	0.82	0.69	0.68	0.67
	LEN	187.26	31.67	7.36	152.34	26.08	6.70	74.37	18.27	5.52
DG REG	RMSE	34.67	7.37	1.92	27.41	6.06	1.90	18.07	4.51	1.66
	MAE	17.87	5.36	1.52	12.21	4.31	1.53	9.45	3.04	1.37
	MRAE	0.61	0.39	0.17	0.77	0.51	0.26	2.35	0.76	0.37
	MAX	245.90	41.32	6.58	226.19	27.97	6.65	189.37	32.60	6.27
	COV	0.90	0.86	0.92	0.77	0.68	0.82	0.60	0.62	0.68
	LEN	102.92	23.76	6.80	68.20	15.92	5.87	43.89	10.47	4.48
Gauss REG	RMSE	34.95	7.43	1.92	28.86	6.06	1.87	16.23	4.43	1.56
	MAE	17.61	5.34	1.52	12.45	4.30	1.50	5.81	2.76	1.20
	MRAE	0.46	0.37	0.17	0.60	0.51	0.25	0.67	0.59	0.31
	MAX	246.62	48.08	6.60	231.78	28.16	6.63	175.54	32.27	6.29
	COV	0.89	0.83	0.88	0.78	0.69	0.79	0.65	0.65	0.67
	LEN	98.95	23.49	6.81	70.33	17.10	6.05	31.58	11.69	4.88
DG CAR	RMSE	30.61	7.50	1.94	23.99	6.60	1.93	8.99	5.94	1.71
	MAE	15.66	5.46	1.54	10.29	4.60	1.56	4.21	3.46	1.39
	MRAE	0.55	0.41	0.18	0.55	0.51	0.27	0.77	0.73	0.38
	MAX	228.87	41.60	6.56	222.10	29.95	6.86	95.55	42.40	6.42
	COV	0.80	0.80	0.91	0.40	0.50	0.80	0.23	0.26	0.67
	LEN	75.80	21.63	6.69	29.02	11.70	5.68	5.32	3.60	4.29
Gauss CAR	RMSE	31.42	7.47	1.93	25.46	6.17	1.90	14.45	4.46	1.59
	MAE	16.00	5.43	1.53	11.20	4.40	1.51	6.16	2.85	1.24
	MRAE	0.50	0.40	0.17	0.65	0.53	0.26	1.18	0.63	0.32
	MAX	232.03	41.54	6.57	218.80	28.20	6.87	165.86	31.98	6.31
	COV	0.82	0.79	0.87	0.70	0.60	0.77	0.67	0.61	0.66
	LEN	80.58	22.04	6.73	57.23	14.99	5.87	34.35	10.50	4.74