Journal of Data Science 24 (1), 187–202
January 2026
DOI: 10.6339/25-JDS1194
Statistical Data Science

Differentially Private Bayesian Envelope Regression via Sufficient Statistic Perturbation

Peng Yu[1,†], Yangdi Jiang[1,†], Zhihua Su[2], Jiamei Wu[3], Lingchen Kong[3], and Bei Jiang[1,*]

[1]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada*
[2]*nVerses Capital, Wellington, FL, USA*
[3]*School of Mathematics and Statistics, Beijing Jiaotong University, Beijing, China*

Abstract

We propose a differentially private Bayesian framework for envelope regression, a technique that improves estimation efficiency by modelling the response as a function of a low-dimensional subspace of the predictors. Our method applies the analytic Gaussian mechanism to privatize sufficient statistics from the data, ensuring formal (ϵ, δ)-differential privacy. We develop a tailored Gibbs sampling algorithm that performs valid Bayesian inference using only the noisy sufficient statistics. This approach leverages the envelope structure to isolate the variation in predictors that is relevant to the response, reducing estimation error compared to standard regression under the same privacy constraints. Through simulation studies, we demonstrate improved estimation accuracy and tighter credible intervals relative to a differentially private Bayesian linear regression baseline.

Keywords *credible interval; dimension reduction; MCMC; statistical inference*

1 Introduction

As technological advancements continue to accelerate, we are faced with the challenge of managing and understanding increasingly complex data. Dimensionality reduction is a fundamental tool for understanding such complex data (Aoshima et al., 2018; Aoshima and Yata, 2017). Despite each data point often comprising many features, the underlying subject of interest is typically lower-dimensional. Reducing the data's "extrinsic" dimension to its "intrinsic" dimension enables analysts to unveil critical structural relationships among features. This dimensionality reduction not only facilitates more efficient utilization of the data for learning tasks, such as classification and regression but also significantly diminishes the storage space required for the data. On the one hand, it streamlines these datasets into a more tractable form, retaining crucial information which facilitates simpler analysis and interpretation. On the other hand, dimensionality reduction techniques are pivotal in diminishing the number of variables under consideration. This reduction is essential for addressing challenges such as the curse of dimensionality and the risk of overfitting in statistical models. By lowering the complexity of the data, these techniques contribute to more robust and generalizable model construction.

*Corresponding author. Email: bei1@ualberta.ca.
†These two authors contributed equally to this paper.

Received December 28, 2024; Accepted June 24, 2025

As data complexity grows, so does the imperative to safeguard data privacy. Differential Privacy (DP) (Dwork et al., 2006b) has gained recognition as a prominent mathematical framework for quantifying privacy protection, and several privacy mechanisms (Dwork et al., 2006a; Yao and Li, 2018; Chanyaswad et al., 2018; Dwork and Roth, 2014) have been devised to achieve DP. It entails the introduction of calibrated random fluctuations into algorithmic calculations to demonstrably constrain the probability of individual-specific information being revealed through the algorithm's output. Such a guarantee protects the privacy of individuals while still allowing valuable insights to be extracted from the data. DP is widely used in various applications, especially in scenarios where sensitive data needs to be analyzed, such as healthcare research (Dyda et al., 2021) and census data analysis (Doe and Roe, 2021). Building upon this foundation, our paper develops an innovative data-driven approach to linear regression with DP by leveraging the envelope concept (Cook et al., 2010). This framework greatly improves the efficiency of coefficient estimation by adeptly filtering out extraneous information among predictors, showcasing a pioneering application of DP in enhancing statistical analysis precision while concurrently securing sensitive data.

Over the past decade, substantial progress has been made in adapting traditional linear regression techniques to comply with differential privacy (DP), with methodologies broadly categorized into frequentist and Bayesian approaches. Frequentist methods, such as sufficient statistic perturbation and subsample aggregation, achieve DP by introducing noise into summary statistics or aggregating results from multiple DP-treated subsamples, as exemplified in the works of Zhang et al. (2016), McSherry and Mironov (2009), and Smith (2008). Bayesian approaches, including MCMC-based data augmentation, integrate MCMC techniques with DP to enable Bayesian linear regression, as demonstrated by Ju et al. (2022) and Bernstein and Sheldon (2019). While much of the existing research focuses on enhancing the accuracy of coefficient estimation under privacy constraints, a smaller subset has explored regression structures beyond utility considerations. Studies such as Dandekar et al. (2018), Chaudhuri et al. (2013), Wang and Xu (2019), and Talwar et al. (2015) incorporate techniques like Principal Component Analysis (PCA) for dimensionality reduction and regularization methods such as Lasso, Ridge, and Sparse Regression to address overfitting and improve interpretability in high-dimensional data. These advancements highlight efforts to balance privacy preservation with maintaining the integrity of regression structures, although the exploration of structural properties in DP linear regression remains an open area for further research and development.

Motivated by the observation that some directions in the predictor space have negligible impact on the response, we propose a differentially private regression framework that separates material and immaterial components of the predictors. This framework is built upon envelope regression, which identifies a low-dimensional subspace capturing all predictive variation. By focusing inference on this subspace, envelope regression improves estimation efficiency, a property especially valuable in the differentially private setting, where noise can otherwise obscure the signal. Our methodology ensures formal (ϵ, δ)-differential privacy by applying the analytic Gaussian mechanism to sufficient statistics of the data. We then develop a Gibbs sampling algorithm for posterior inference that respects both the envelope structure and the privatized nature of the statistics. This algorithm is a specialization of the general data augmentation MCMC framework proposed by Ju et al. (2022), tailored specifically to the envelope regression model. While Ju et al. (2022) derived their sampler for standard linear and log-linear models, our contribution extends this approach to a structured regression setting that leverages dimensionality reduction for improved utility. The remainder of the paper outlines the theoretical background, details the

sampling algorithm, and presents simulation results comparing our method to a differentially private Bayesian linear regression baseline.

2 Preliminaries

In this section, we revisit some background material for differential privacy (Dwork et al., 2006b) and envelope methodology (Cook et al., 2010).

2.1 Differential Privacy

Differential privacy is a privacy definition that is tailored to the task of privacy-preserving data analysis. First introduced in Dwork et al. (2006b), it quickly gained popularity as it provides a mathematically rigorous framework to quantify the amount of privacy protection. The formal definition of ϵ-Differential Privacy is given as follows.

Definition 1 (ϵ-Differential Privacy (Dwork et al., 2006b))**.** For any $\epsilon \geqslant 0$, a mechanism $\mathcal{A}$ is said to be ϵ-differentially private (ϵ-DP) if for all measurable sets $\mathcal{S}$ and for all pairs of neighbouring datasets $\mathbf{X}$ and $\mathbf{X}'$, where neighbouring datasets refer to two datasets that differ by only one element, the following holds,

$$\mathbb{P}\big(\mathcal{A}(\mathbf{X}) \in \mathcal{S}\big) \leqslant \exp(\epsilon)\mathbb{P}\big(\mathcal{A}\big(\mathbf{X}'\big) \in \mathcal{S}\big). \tag{1}$$

The parameter ϵ quantifies the level of privacy protection. Smaller values of ϵ provide stronger privacy guarantees. When ϵ is zero, there is no privacy protection, and as ϵ increases, the privacy protection decreases. A natural relaxation of ϵ-DP is the (ϵ, δ)-differential privacy. It has found widespread application in practical settings where a small amount of privacy leakage is acceptable in exchange for improved accuracy or utility of the analysis. In this paper, we will focus on (ϵ, δ)-differential privacy as our privacy definition.

Definition 2 ((ϵ, δ)-Differential Privacy (Dwork et al., 2006a,b))**.** For any $\delta \in [0, 1]$ and $\epsilon \geqslant 0$, a mechanism $\mathcal{A}$ is said to be (ϵ, δ)-differentially private if for all measurable sets $\mathcal{S}$ and for all pairs of neighboring datasets $\mathbf{X}$ and $\mathbf{X}'$, the following holds true,

$$\mathbb{P}\big(\mathcal{A}(\mathbf{X}) \in \mathcal{S}\big) \leqslant \exp(\epsilon)\mathbb{P}\big(\mathcal{A}\big(\mathbf{X}'\big) \in \mathcal{S}\big) + \delta. \tag{2}$$

To achieve (ϵ, δ)-DP, the analytic Gaussian mechanism is one of the most widely used mechanisms (Balle and Wang, 2018). It improves the original Gaussian mechanism by calibrating the variance directly using the Gaussian cumulative distribution function instead of a tail-bound approximation. Before introducing the mechanism, we will need the following definition.

Definition 3 (l_2-sensitivity (Dwork et al., 2006a))**.** Let $\mathcal{D}$ denote the space of the confidential datasets. The l_2-sensitivity of a query function $f : \mathcal{D} \to \mathbb{R}^d$ is defined as $\Delta_2 = \max_{\mathbf{X},\mathbf{X}'} \|f(\mathbf{X}) - f(\mathbf{X}')\|_2$ for all pairs of neighboring datasets $\mathbf{X}$ and $\mathbf{X}'$.

Intuitively, the larger the sensitivity, the larger the noise that needs to be injected into the query, resulting in the same level of privacy protection. Using this notion of sensitivity, we can formally introduce the analytic Gaussian mechanism.

Definition 4 (Analytic Gaussian Mechanism (Balle and Wang, 2018))**.** Let $f : \mathcal{D} \to \mathbb{R}^d$ be a function with l_2 sensitivity Δ_2. For any $\epsilon \geqslant 0$ and $\delta \in [0, 1]$, the Gaussian output perturbation mechanism $M(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma_{\text{dp}}^2 \text{I})$ is (ϵ, δ)-differentially private if and only if

$$\Phi\left(\frac{\Delta_2}{2\sigma_{\text{dp}}} - \frac{\epsilon\sigma_{\text{dp}}}{\Delta_2}\right) - e^{\epsilon}\Phi\left(-\frac{\Delta_2}{2\sigma_{\text{dp}}} - \frac{\epsilon\sigma_{\text{dp}}}{\Delta_2}\right) \leqslant \delta, \tag{3}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of standard normal distribution.

To find the noise variance σ_{dp}^2 that satisfies the inequality (3), the numeric algorithm in Balle and Wang (2018) can be used.

2.2 Predictor Envelope Regression

Consider the following univariate linear regression model with p predictor variables,

$$y_i = \boldsymbol{x}_i^{\top}\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \overset{i.i.d}{\sim} \mathcal{N}\big(0, \sigma^2\big), \quad i = 1, \ldots, n, \tag{4}$$

where y_i represents the scalar response of the i-th observation, and $\boldsymbol{x}_i \in \mathbb{R}^p$ represents the predictor vector of the i-th observation with a mean of $\mathbf{0}$ and a covariance matrix of $\boldsymbol{\Sigma}_{\mathbf{x}}$. In many applications, only a subset of the variation in $\boldsymbol{x}_i$ influences y_i; the remaining variation introduces noise and reduces estimation efficiency.

The envelope approach (Cook and Zhang, 2015; Cook et al., 2010) seeks a subspace of the predictor space that captures all the variation in $\boldsymbol{x}_i$ that is relevant to predicting y_i. Formally, it partitions the predictor space into a material component (which affects the response) and an immaterial component (which does not), and then estimates the regression coefficients using only the material component. Let $\mathcal{E} \subseteq \mathbb{R}^p$ be the smallest subspace satisfying the following:

1. The projection of $\boldsymbol{x}_i$ onto the orthogonal complement of $\mathcal{E}$ is uncorrelated with its projection onto $\mathcal{E}$.
2. Given the projection onto $\mathcal{E}$, the response y_i is uncorrelated with the projection of $\boldsymbol{x}_i$ onto the orthogonal complement.

Under this structure, the regression depends only on the projection of $\boldsymbol{x}_i$ onto $\mathcal{E}$, which implies that the response model can be expressed as:

$$y_i = \boldsymbol{x}_i^{\top}\mathbf{B}_1\boldsymbol{\theta} + \varepsilon_i,$$

where the columns of $\mathbf{B}_1 \in \mathbb{R}^{p\times r}$ form an orthonormal basis for $\mathcal{E}$, and $\boldsymbol{\theta} \in \mathbb{R}^r$ is the reduced-dimensional coefficient vector with $\boldsymbol{\beta} = \mathbf{B}_1\boldsymbol{\theta}$. The covariance of $\boldsymbol{x}_i$ is then decomposed into two orthogonal components associated with the material and immaterial parts:

$$\Sigma_x = \mathbf{B}_1\Omega\mathbf{B}_1^{\top} + \mathbf{B}_2\Omega_0\mathbf{B}_2^{\top},$$

where $\mathbf{B}_2$ spans the immaterial subspace orthogonal to $\mathcal{E}$.

In essence, envelope regression estimates the model parameters by focusing only on the subspace of the predictors that is informative for the response, leading to potentially large gains in efficiency, especially in high-dimensional settings or when predictor variability is not uniformly relevant.

3 Data Augmentation MCMC via Privatized Sufficient Statistics

In this section, we introduce a privacy-preserving, data-driven framework for Bayesian linear regression. The proposed methodology facilitates valid differentially private Bayesian inference by leveraging sufficient statistics perturbed through the analytic Gaussian mechanism. A key innovation of our approach lies in its ability to account for the low-dimensional structure of the data by partitioning predictors into material and immaterial components. This partitioning is informed by the observation that, in certain real-world applications, specific variations in some predictors may exert no discernible influence on the response variable.

Data Augmentation Denote $\boldsymbol{x}_i$ as the i-th p-dimensional predictor, y_i as the corresponding scalar response, and $\boldsymbol{\nu} = (\mathbf{B}_1, \mathbf{B}_2, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\theta}, \sigma)$ as the collection of model parameters. Consider the confidential dataset represented by $\mathbf{x} \doteq (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top \in \mathbb{R}^{n \times p}$ and $\mathbf{y} \doteq (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$. Instead of having direct access to the confidential dataset $(\mathbf{x}, \mathbf{y})$, our observations are limited to the privatized sufficient statistics for $\boldsymbol{\nu}$ which we denote as $\boldsymbol{s}_{\text{dp}}$. Under the Bayesian approach, we are concerned with the following posterior distribution:

$$p(\boldsymbol{\nu} \mid \boldsymbol{s}_{\text{dp}}) \propto p(\boldsymbol{\nu}) p(\boldsymbol{s}_{\text{dp}} \mid \boldsymbol{\nu}). \tag{5}$$

As the marginal likelihood $p(\boldsymbol{s}_{\text{dp}} \mid \boldsymbol{\nu})$ is often unknown, we augment the MCMC state space with the latent confidential dataset $(\mathbf{x}, \mathbf{y})$,

$$p(\boldsymbol{\nu}, \mathbf{x}, \mathbf{y} \mid \boldsymbol{s}_{\text{dp}}) \propto p(\boldsymbol{\nu}) f(\mathbf{x}, \mathbf{y} \mid \boldsymbol{\nu}) p(\boldsymbol{s}_{\text{dp}} \mid \mathbf{x}, \mathbf{y}). \tag{6}$$

Gibbs Sampler Marginally, the $\boldsymbol{\nu}$ samples produced by equation (6) follow the posterior $p(\boldsymbol{\nu} \mid \boldsymbol{s}_{\text{dp}})$ in equation (5). The joint posterior distribution of $p(\boldsymbol{\nu}, \mathbf{x}, \mathbf{y} \mid \boldsymbol{s}_{\text{dp}})$ can be achieved through the following Gibbs sampling procedure: (a) sample the confidential dataset $(\mathbf{x}, \mathbf{y})$ given model parameters $\boldsymbol{\nu}$; (b) sample parameters $\boldsymbol{\nu}$ given latent confidential $(\mathbf{x}, \mathbf{y})$ and $\boldsymbol{s}_{\text{dp}}$ (Ju et al., 2022). The following subsections will illustrate the envelope regression structure $[\mathbf{x}, \mathbf{y} | \boldsymbol{\nu}]$, the sufficient statistics $[\boldsymbol{s}_{\text{dp}} \mid \mathbf{x}, \mathbf{y}]$ and their corresponding Gibbs sampling procedures.

3.1 Hierarchical Envelope Linear Regression

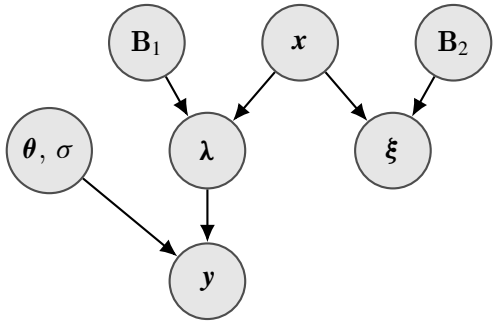


Figure 1: DAG of the envelope regression model.

The hierarchical envelope regression structure between $(\mathbf{x}, \mathbf{y})$ given parameters $\boldsymbol{\nu} = (\mathbf{B}_1, \mathbf{B}_2, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\theta}, \sigma)$ can be expressed as follows,

$$\begin{aligned} \boldsymbol{x}_i &= \mathbf{B}_1\boldsymbol{\lambda}_i + \mathbf{B}_2\boldsymbol{\xi}_i, \quad \boldsymbol{\lambda}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \quad \boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_0), \\ y_i &= \boldsymbol{\lambda}_i^\top \boldsymbol{\theta} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\big(0, \sigma^2\big). \end{aligned} \tag{7}$$

In model (7), $\mathbf{B}_1 \in \mathbb{R}^{p\times r}$ and $\mathbf{B}_2 \in \mathbb{R}^{p\times(p-r)}$ are two orthogonal matrices with $\mathbf{B}_1^\top\mathbf{B}_2 = \mathbf{0}$. It can be seen that $\boldsymbol{x}_i$ is decomposed into two parts, which are the material part $\mathbf{B}_1\boldsymbol{\lambda}_i$ and the immaterial part $\mathbf{B}_2\boldsymbol{\xi}_i$. Note that $\boldsymbol{\lambda}_i \in \mathbb{R}^r$ and $\boldsymbol{\xi}_i \in \mathbb{R}^{p-r}$ are the corresponding coordinates of material and immaterial. The regression dependency between $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ is only mediated by the material part $\boldsymbol{\lambda}_i$. After dimension reduction, the regression coefficient $\boldsymbol{\theta}$ is r-dimensional. See Figure 1 for an illustration of the hierarchical envelope regression structure.

Identifiability of $\mathbf{B}_1$ and $\mathbf{B}_2$ As shown in model (7), y_i remains the same no matter how $\boldsymbol{x}_i$ varies in $\text{span}(\mathbf{B}_2)$ since the distribution of $[y_i|\boldsymbol{x}_i]$ is the same as $[y_i|\mathbf{B}_1^\top\boldsymbol{x}_i]$. It is worth noting that for the given dimension r, $\mathbf{B}_1$ and $\mathbf{B}_2$ are not uniquely defined. To ensure the identifiability, we define $\mathbf{B}_1$ and $\mathbf{B}_2$ as a function of an unconstrained matrix $\mathbf{A} \in \mathbb{R}^{(p-r)\times r}$. Let $\mathbf{C}_\mathbf{A} = (\mathbf{I}_r, \mathbf{A}^\top)^\top$ and $\mathbf{D}_\mathbf{A} = (-\mathbf{A}, \mathbf{I}_{p-r})^\top$, define

$$\mathbf{B}_1(\mathbf{A}) \doteq \mathbf{C}_\mathbf{A}\big(\mathbf{C}_\mathbf{A}{}^\top\mathbf{C}_\mathbf{A}\big)^{-1/2}, \quad \mathbf{B}_2(\mathbf{A}) \doteq \mathbf{D}_\mathbf{A}\big(\mathbf{D}_\mathbf{A}{}^\top\mathbf{D}_\mathbf{A}\big)^{-1/2}. \tag{8}$$

In equation (8), the matrix $\mathbf{A}$ and $\text{span}(\mathbf{B}_1)$ are uniquely determined by each other, and hence $\mathbf{B}_1(\mathbf{A})$ is identifiable. The notation of $\mathbf{B}_1(\mathbf{A})$ as the function of $\mathbf{A}$ prevents the prior selection and the posterior updating of $\mathbf{B}_1$ on Stiefel manifolds. As there is no close form posterior for $\mathbf{A}$, the update of $\mathbf{A}$ is through the Metropolis-Hastings algorithm, see Appendix A.1 for more details.

Likelihood of $[\mathbf{x}, \mathbf{y}|\boldsymbol{\nu}]$ Given model (7), the conditional distribution of $[y_i|\boldsymbol{x}_i]$ is uniquely determined with $[y_i|\boldsymbol{x}_i, \mathbf{B}_1, \boldsymbol{\theta}] \sim \mathcal{N}(\boldsymbol{x}_i^\top\mathbf{B}_1\boldsymbol{\theta}, \sigma^2)$, and the regression coefficient $\boldsymbol{\beta}$ of $[y_i|\boldsymbol{x}_i]$ is given by $\boldsymbol{\beta} = \mathbf{B}_1\boldsymbol{\theta}$. The density function $f(\mathbf{x}, \mathbf{y}|\boldsymbol{\nu})$ is as follows,

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}|\boldsymbol{\nu}) &= f\big(\mathbf{y}|\mathbf{x}, \mathbf{B}_1, \boldsymbol{\theta}, \sigma^2\big) f(\mathbf{x}|\mathbf{B}_1, \mathbf{B}_2, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0) \\ &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y_i - \boldsymbol{x}_i^\top\mathbf{B}_1\boldsymbol{\theta})^2}{2\sigma^2}\right] \times \prod_{i=1}^n \frac{1}{|\boldsymbol{\Sigma}_\mathbf{x}|^{1/2}2\pi^{\,p/2}} \exp\left[-\frac{\boldsymbol{x}_i^\top\boldsymbol{\Sigma}_\mathbf{x}^{-1}\boldsymbol{x}_i}{2}\right], \end{aligned} \tag{9}$$

where $\boldsymbol{\Sigma}_\mathbf{x} = \mathbf{B}_1\boldsymbol{\Omega}\mathbf{B}_1^\top + \mathbf{B}_2\boldsymbol{\Omega}_0\mathbf{B}_2^\top$.

3.2 Differentially Privatized Sufficient Statistic

In model (7), $(\mathbf{x}^\top\mathbf{x}, \mathbf{x}^\top\mathbf{y}, \mathbf{y}^\top\mathbf{y})$ is the sufficient statistics of $\boldsymbol{\nu} = (\mathbf{B}_1, \mathbf{B}_2, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\theta}, \sigma)$. Denote the l_2 sensitivity for $(\mathbf{x}^\top\mathbf{x}, \mathbf{x}^\top\mathbf{y}, \mathbf{y}^\top\mathbf{y})$ as Δ_2. Using the analytic Gaussian mechanism, we generate the differentially private sufficient statistics $\boldsymbol{s}_{\text{dp}}$ by adding Gaussian noises $\mathcal{N}(0, \sigma_{\text{dp}}^2)$ independently to each element of $(\mathbf{x}^\top\mathbf{x}, \mathbf{x}^\top\mathbf{y}, \mathbf{y}^\top\mathbf{y})$ where σ_{dp}^2 satisfies the inequality (3).

Sensitivity of the Sufficient Statistics Note that to apply the analytic Gaussian mechanism, the sensitivity Δ_2 must be finite. In differential privacy literature, the routine procedure entails bounding each predictor and response variable in a manner that is independent of the

data (Bernstein and Sheldon, 2018, Section 3.3). For simplicity, we assume a lower and upper bound (L, U) for all dimensions of $\boldsymbol{x}_i$ and y_i. To compute Δ_2, we can reason the worst case influence of an individual on each component of $\mathbf{s}_{\text{dp}} = (\mathbf{x}^\top\mathbf{x}, \mathbf{x}^\top\mathbf{y}, \mathbf{y}^\top\mathbf{y})$. The number of unique elements in $\mathbf{s}_{\text{dp}}$ is $(p(p+1)/2, p, 1)$ respectively, and we have

$$\Delta_2 = (U-L)^2 p(p+1)/2 + (U-L)^2(p+1). \tag{10}$$

3.3 Prior Specification

Recall that we denote the parameters in the model (7) as $\boldsymbol{\nu} = (\mathbf{B}_1, \mathbf{B}_2, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\theta}, \sigma)$ and that $\mathbf{B}_1$, $\mathbf{B}_2$ are set to be functions of an unconstrained matrix $\mathbf{A}$ for identification. Thus, we impose diffuse priors for $(\mathbf{A}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\theta}, \sigma)$ as follows,

- $\boldsymbol{\theta}$ is normally distributed: $\boldsymbol{\theta} \sim \mathcal{N}(0, 100\mathbf{I}_r)$.
- σ^2 follows Inverse-Gamma distribution: $\sigma^2 \sim \text{IG}(a_0, b_0)$ where $a_0 = b_0 = 0.1$.
- $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ follow Inverse-Wishart distribution: $\boldsymbol{\Omega} \sim \text{IW}(\mathbf{S}, \text{s})$, $\boldsymbol{\Omega}_0 \sim \text{IW}(\mathbf{S}_0, \text{s}_0)$, where the parameters are commonly selected as $\mathbf{S} = \mathbf{I}_{p-r}$, $\mathbf{S}_0 = \mathbf{I}_r$ and $\text{s} = p - r + 1$, $\text{s}_0 = r + 1$ (Frühwirth-Schnatter, 2006).
- $\mathbf{A}$ follows matrix normal distribution: $\mathbf{A} \sim \mathcal{MN}(\mathbf{A}_0, \mathbf{K}, \mathbf{L})$ where $\mathbf{A}_0 \in \mathbb{R}^{(p-r)\times r}$ is the mean matrix for $\mathbf{A}$, and $\mathbf{K} \in \mathbb{S}^{(p-r)\times(p-r)}$, $\mathbf{L} \in \mathbb{S}^{r\times r}$ are symmetric positive definite covariance matrices. Here we set $\mathbf{A}_0 = \mathbf{0}$, which indicates that $\mathbf{B}_1$ is assumed to be centered at $\mathbf{B}_1(\mathbf{0}) = (\mathbf{I}_r, \mathbf{0})^\top$, which corresponds to an identity projection to the first r dimensions of the predictors. The covariance matrices $\mathbf{K}$ and $\mathbf{L}$ are chosen to be $10\mathbf{I}_{(p-r)}$ and $10\mathbf{I}_r$.

3.4 Privacy-Aware Gibbs Sampler

Let $\boldsymbol{\lambda} \doteq (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n)^\top$ and $\boldsymbol{\xi} \doteq (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)^\top$ be the material and immaterial information for the confidential dataset. Let $(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\xi}^{(t)}, \boldsymbol{\nu}^{(t)})$ denote the state of the Gibbs sampler at the t-th iterations. Based on equation (6), the Gibbs sampling procedure can be split into three detailed steps as follows.

1. **Sample $(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)})$ given $\boldsymbol{\nu}^{(t)}$ and $\boldsymbol{s}_{\text{dp}}$**: the conditional distribution of $(\mathbf{x}, \mathbf{y})$ given $\boldsymbol{s}_{\text{dp}}$ and $\boldsymbol{\nu}^{(t)}$ has no closed form posterior. Here we employ the algorithm in Ju et al. (2022) for the sampler which we state in Algorithm 2. Note that the notation $\mathcal{N}(\boldsymbol{s}_{\text{dp}}; \boldsymbol{t}_s, \sigma_{\text{dp}}^2\mathbf{I})$ represents the probability density of $\mathcal{N}(\boldsymbol{t}_s, \sigma_{\text{dp}}^2\mathbf{I})$ in $\boldsymbol{s}_{\text{dp}}$.
2. **Calculate $(\boldsymbol{\lambda}^{(t+1)}, \boldsymbol{\xi}^{(t+1)})$ based on $\mathbf{x}^{(t+1)}$ and $\boldsymbol{\nu}^{(t)}$**:

$$\boldsymbol{\lambda}^{(t+1)} = \mathbf{x}^{(t+1)}\mathbf{B}_1^{(t)\top}, \quad \boldsymbol{\xi}^{(t+1)} = \mathbf{x}^{(t+1)}\mathbf{B}_2^{(t)\top}.$$

3. **Sample $\boldsymbol{\nu}^{(t+1)}$ given $(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)})$**: to obtain the $(t+1)$-th iteration of the parameter $\boldsymbol{\nu}$, we refer the reader to Appendix A.1 as it is a bit too long to be included here.

This 3-step procedure is summarized in Algorithm 1.

4 Simulation Studies

In this section, we conduct simulation studies to evaluate the performance of our proposed MCMC framework for Bayesian envelope linear regression under differential privacy. The goal is to assess whether leveraging the envelope structure enhances the efficiency of coefficient estimation and uncertainty quantification compared to a baseline method that does not account

Algorithm 1 Privacy-Aware Gibbs sampler.

Input: current state of the Gibbs sampler $(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{\xi}^{(t)}, \boldsymbol{\nu}^{(t)})$, DP sufficient statistics $\boldsymbol{s}_{\text{dp}}$.
Output: next state of the Gibbs sampler $(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \boldsymbol{\lambda}^{(t+1)}, \boldsymbol{\xi}^{(t+1)}, \boldsymbol{\nu}^{(t+1)})$.

1: **Sample** $(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)})$ given $\boldsymbol{\nu}^{(t)}$ and $\boldsymbol{s}_{\text{dp}}$ using Algorithm 2.
2: **Compute** $\boldsymbol{\lambda}^{(t+1)} = \mathbf{x}^{(t+1)}\mathbf{B}_1^{(t)\top}$ and $\boldsymbol{\xi}^{(t+1)} = \mathbf{x}^{(t+1)}\mathbf{B}_2^{(t)\top}$.
3: **Sample** $\boldsymbol{\nu}^{(t+1)}$ given $(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \boldsymbol{\lambda}^{(t+1)}, \boldsymbol{\xi}^{(t+1)})$ using the procedure detailed in Appendix A.1.
4: **Return:** $(\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \boldsymbol{\lambda}^{(t+1)}, \boldsymbol{\xi}^{(t+1)}, \boldsymbol{\nu}^{(t+1)})$.

Algorithm 2 Update $(\mathbf{x}, \mathbf{y})$ within Gibbs sampler.

Input: current state $(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\xi}, \boldsymbol{\nu})$, DP sufficient statistics $\boldsymbol{s}_{\text{dp}}$.
Output: new state $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$.

Compute $\boldsymbol{t}_s = \{\mathbf{x}^\top\mathbf{y}, \mathbf{x}^\top\mathbf{x}, \mathbf{y}^\top\mathbf{y}\}$ with the current $(\mathbf{x}, \mathbf{y})$.
for each $i \in [n]$ **do**
  Propose new $(\boldsymbol{x}_i^\star, y_i^\star)$ as follows:
    $\boldsymbol{x}_i^\star \sim \mathcal{N}(\mathbf{0}, \mathbf{B}_1\boldsymbol{\Omega}\boldsymbol{B}^\top + \mathbf{B}_2\boldsymbol{\Omega}_0\mathbf{B}_2^\top)$
    $y_i^\star|\boldsymbol{x}_i^\star \sim \mathcal{N}(\boldsymbol{x}_i^{\star\top}\mathbf{B}_1\boldsymbol{\theta}, \sigma^2)$
  Compute $\boldsymbol{t}_s^+ = \boldsymbol{t}_s - \boldsymbol{t}_i + \boldsymbol{t}_i^\star$ where
    $\boldsymbol{t}_i = \{\boldsymbol{x}_i y_i, \boldsymbol{x}_i\boldsymbol{x}_i^\top, y_i^2\}$,
    $\boldsymbol{t}_i^\star = \{\boldsymbol{x}_i^\star y_i^\star, \boldsymbol{x}_i^\star\boldsymbol{x}_i^{\star\top}, y_i^{\star 2}\}$.
  Accept the proposed state with probability $\alpha(\boldsymbol{x}_i^\star, y_i^\star)$ given by:
    $\alpha(\boldsymbol{x}_i^\star, y_i^\star) = \min\left\{\frac{\mathcal{N}(\boldsymbol{s}_{\text{dp}}; \boldsymbol{t}_s^+, \sigma_{\text{dp}}^2\mathbf{I})}{\mathcal{N}(\boldsymbol{s}_{\text{dp}}; \boldsymbol{t}_s, \sigma_{\text{dp}}^2\mathbf{I})}, 1\right\}$
  Set $\boldsymbol{t}_s = \boldsymbol{t}_s^+$ and $(\tilde{\boldsymbol{x}}_i, \tilde{y}_i) = (\boldsymbol{x}_i^\star, y_i^\star)$ if the state is accepted. Otherwise, set $(\tilde{\boldsymbol{x}}_i, \tilde{y}_i) = (\boldsymbol{x}_i, y_i)$.
end for
Return $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = (\tilde{\boldsymbol{x}}_i, \tilde{y}_i)_{i=1}^n$.

for this structure. We compare our approach against the data augmentation MCMC framework proposed by Ju et al. (2022) for Bayesian inference under differential privacy, which includes linear regression as one of its settings but does not incorporate envelope structure. Our comparison focuses on whether embedding envelope structure leads to lower mean squared error (MSE) and narrower credible intervals under privacy constraints. Additionally, we examine how these performance metrics vary with the privacy budget ϵ and sample size n.

Throughout the experiments, the predictor dimension is fixed at $p = 4$, and the envelope dimension is set to $r = 2$. The privacy budget ϵ is varied across $\{0.5, 1, 3, 5\}$ while δ is set to $1/n$. Sample sizes considered are $n \in \{500, 1000, 5000\}$. For each setting, we generate 50 independent datasets and compute differentially private sufficient statistics using the analytic Gaussian mechanism.

4.1 Generating the Differentially Private Sufficient Statistics

We generate synthetic datasets according to the envelope regression model described in Section 3, with details outlined in Algorithm 3.

We fix the total predictor dimension to $p = 4$ and the envelope dimension to $r = 2$, so that half of the predictor variability is immaterial to the response. The sample size varies across $n \in \{500, 1000, 5000\}$ to represent low-, moderate-, and high-data regimes. To assess the impact

Algorithm 3 Generating simulation datasets.

Input: privacy budget ϵ, δ, covariance matrices $\mathbf{\Omega} \in \mathbb{R}^{r\times r}$, $\mathbf{\Omega_0} \in \mathbb{R}^{(p-r)\times(p-r)}$, unconstrained matrix $\mathbf{A} \in \mathbb{R}^{(p-r)\times r}$, clamping interval $[L, U]$, regression coefficients $\boldsymbol{\theta} \in \mathbb{R}^r$, error variance σ^2 and sample size n.

Output: clamped dataset $(\boldsymbol{x}_i^C, y_i^C)_{i=1}^n$

Compute compute $\mathbf{B}_1$ and $\mathbf{B}_2$ using $\mathbf{A}$ through (8).
for each $i \in [n]$ **do**
  Generate $\boldsymbol{\lambda}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$ and $\boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_0)$.
  Compute $\boldsymbol{x}_i = \mathbf{B}_1\boldsymbol{\lambda}_i + \mathbf{B}_2\boldsymbol{\xi}_i$.
  Generate $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.
  Compute $y_i = \boldsymbol{\lambda}_i^\top\boldsymbol{\theta} + \varepsilon_i$.
  Clamp all entries of $\boldsymbol{x}_i$ and y_i within the interval (L, U). Denote them as $\boldsymbol{x}_i^C$ and y_i^C.
end for
Return: $(\boldsymbol{x}_i^C, y_i^C)_{i=1}^n$.

of privacy constraints, we vary the privacy budget over $\epsilon \in \{0.5, 1, 3, 5\}$, with $\delta = 1/n$ fixed throughout.

For each dataset, we sample the unconstrained matrix $\mathbf{A} = (a_{ij})$ as $a_{ij} \sim \mathcal{N}(5, 3^2)$. The material covariance is fixed as $\mathbf{\Omega} = \mathbf{I}_r$, and the immaterial covariance is set to $\mathbf{\Omega}_0 = 0.1\mathbf{I}_{(p-r)}$, representing weaker variation in the irrelevant subspace. We fix the regression coefficient $\boldsymbol{\theta} = (2, -2)^\top$ and the error variance $\sigma^2 = 1$.

To enforce bounded sensitivity for private mechanisms, we set a bound $[L, U]$ for all dimensions of $\boldsymbol{x}_i$ and y_i by clamping them within the interval $[L, U]$. For a real value z, and $L \leqslant U$, we define the clamp function $[z]_L^U := \min\{\max\{z, L\}, U\}$. If z is a vector of length d, we use the same notation to apply an entry-wise clamp: $[z]_L^U := ([z_1]_L^U, [z_2]_L^U, \ldots, [z_d]_L^U)^\top$. The lower and upper bound is set to be $[-10, 10]$. Once the clamped dataset $(\mathbf{x}_C, \mathbf{y}_C)$ is generated, we generate the differentially private sufficient statistics $\boldsymbol{s}_{\text{dp}}$ through the analytic Gaussian mechanism by adding Gaussian noises independently to each element of $(\mathbf{x}_C^\top\mathbf{x}_C, \mathbf{x}_C^\top\mathbf{y}_C, \mathbf{y}_C^\top\mathbf{y}_C)$. Please refer to Section 3.2 for more details.

For each combination of n and ϵ, we simulate 50 datasets $(\mathbf{x}_C, \mathbf{y}_C)$ and $\boldsymbol{s}_{\text{dp}}$ as described above. For each generated $\boldsymbol{s}_{\text{dp}}$, we obtain the posterior samples of all model parameters and latent variables using the Gibbs sampling algorithm described in Algorithm 1, retaining 5000 iterations after a burn-in period of 5000 iterations.

Ergodicity of the MCMC Procedure An essential aspect of MCMC is ergodicity (Tierney, 1994), which guarantees the convergence of the MCMC chain to the posterior distribution in terms of total variation. It can be verified that within our model: (a) the chosen priors are proper and $p(\boldsymbol{\nu}) > 0$ for all $\boldsymbol{\nu}$; (b) the condition $f(\mathbf{x}, \mathbf{y}|\boldsymbol{\nu}) > 0$ and $p(\boldsymbol{s}_{\text{dp}}|\mathbf{x}, \mathbf{y}) > 0$ is consistently satisfied for all $(\mathbf{x}, \mathbf{y})$. Under the two conditions above, it can be proved that the Gibbs sampler for latent confidential dataset $(\mathbf{x}, \mathbf{y})$ and parameters $\boldsymbol{\nu}$ is ergodic and the limiting distribution is unique (Ju et al., 2022).

4.2 Evaluation and Comparison

As mentioned, we compare our approach against a baseline model: the linear regression setting within the data augmentation MCMC framework proposed by Ju et al. (2022). However, the

baseline model treats the full coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^p$ as the parameter of interest. In contrast, our method adopts the envelope regression framework, where the parameter of interest is the lower-dimensional coefficient $\boldsymbol{\theta} \in \mathbb{R}^r$, associated with the material subspace. However, because our model estimates both the envelope basis matrix $\mathbf{B}_1$ and $\boldsymbol{\theta}$, we are able to recover an estimate of the full coefficient vector via the relation $\boldsymbol{\beta} = \mathbf{B}_1\boldsymbol{\theta}$. In particular, during Gibbs sampling, the t-th posterior draw of $\mathbf{B}_1$ and $\boldsymbol{\theta}$ is denoted as $\hat{\mathbf{B}}_1^{(t)}$ and $\hat{\boldsymbol{\theta}}^{(t)}$, and $\hat{\boldsymbol{\beta}}^{(t)}$ is calculated as $\hat{\boldsymbol{\beta}}^{(t)} = \hat{\mathbf{B}}_1^{(t)}\hat{\boldsymbol{\theta}}^{(t)}$.

Evaluation Metrics We evaluate performance using two metrics: mean squared error (MSE) of the posterior mean estimator and the average width of 95% credible intervals based on the 50 simulated confidential dataset $(\mathbf{x}, \mathbf{y})$ and $\boldsymbol{s}_{\text{dp}}$ in each setup. These capture two core aspects of private Bayesian inference: estimation accuracy and uncertainty quantification.

Denote $\boldsymbol{\beta}^j$ as the true coefficient in the jth simulation and $\hat{\boldsymbol{\beta}}^j$ as its posterior mean estimate, we define MSE as follows,

$$\text{MSE} = \frac{1}{50}\sum_{j=1}^{50}\left\{\frac{1}{p}\left\|\hat{\boldsymbol{\beta}}^j - \boldsymbol{\beta}^j\right\|_2^2\right\},$$

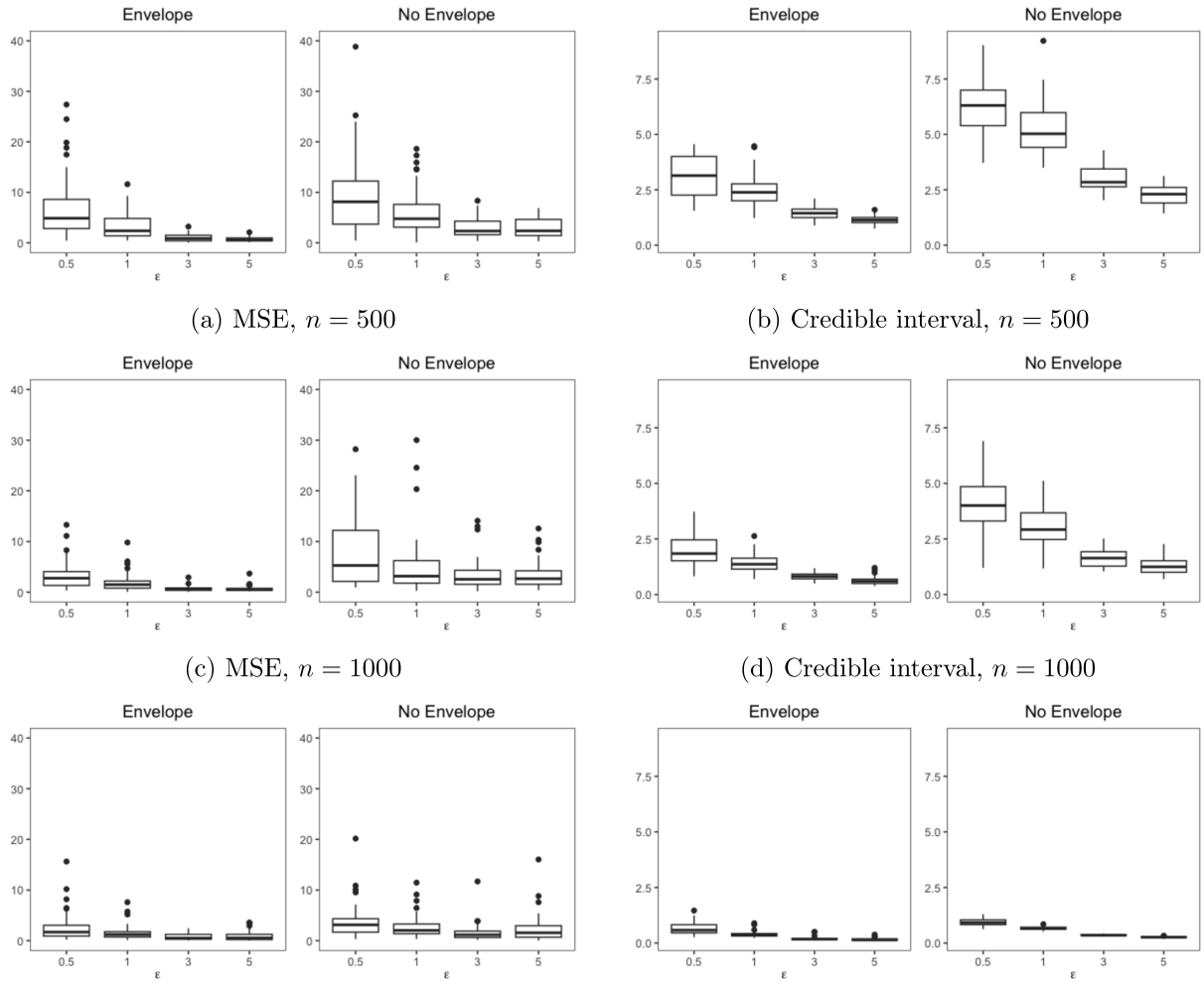
where $\|\cdot\|_2$ represents the L_2 norm. Similarly, the overall average 95% credible interval width is then defined as

$$\text{W} = \frac{1}{50}\sum_{j=1}^{50}\left\{\frac{1}{p}\|\text{W}_{\hat{\boldsymbol{\beta}}^j}\|_2^2\right\},$$

where $\text{W}_{\hat{\boldsymbol{\beta}}^j} \doteq (\text{w}_{\hat{\beta}_1^j}, \ldots, \text{w}_{\hat{\beta}_p^j})$ is the 95% credible interval width vector for all of the p predictors in the jth simulation.

Table 1: The average MSE and interval width for $\hat{\boldsymbol{\beta}}$, derived from our framework utilizing the envelope technique, alongside the framework without the envelope approach.

		MSE		Interval width	
n	ϵ	Bayes Env	No Env	Bayes Env	No Env
$n = 500$	$\epsilon = 0.5$	6.81	11.20	3.10	6.36
	$\epsilon = 1.0$	3.26	6.26	2.48	5.22
	$\epsilon = 3.0$	1.05	3.01	1.45	2.98
	$\epsilon = 5.0$	0.71	2.98	1.14	2.27
$n = 1000$	$\epsilon = 0.5$	3.30	7.56	2.00	4.04
	$\epsilon = 1.0$	1.93	5.19	1.43	3.01
	$\epsilon = 3.0$	0.67	3.42	0.82	1.65
	$\epsilon = 5.0$	0.65	3.34	0.62	1.28
$n = 5000$	$\epsilon = 0.5$	2.67	3.97	0.65	0.94
	$\epsilon = 1.0$	1.60	2.14	0.39	0.66
	$\epsilon = 3.0$	0.81	1.07	0.20	0.36
	$\epsilon = 5.0$	0.91	0.95	0.16	0.27



(a) MSE, $n = 500$

(b) Credible interval, $n = 500$

(c) MSE, $n = 1000$

(d) Credible interval, $n = 1000$

(e) MSE, $n = 5000$

(f) Credible interval, $n = 5000$

Figure 2: The boxplots of average MSE and 95% credible interval width for the coefficient estimation based on our framework and the framework without the envelope approach.

Simulation Results Table 1 reports the average MSE and credible interval width for our envelope-based method compared with the baseline across different values of n and ϵ. Across all configurations, our method consistently achieves lower MSE and narrower credible intervals, with the most pronounced improvements observed in low-sample and high-privacy regimes (e.g., $n = 500$, $\epsilon = 0.5$). These gains stem from the model's ability to isolate signal-relevant variation in the predictors, which helps mitigate the impact of privacy-induced noise on estimation.

As both n and ϵ increase, the performance gap between the two methods narrows. In high-data or low-privacy settings, the baseline model recovers more information from the privatized statistics, reducing the relative benefit of envelope structure. However, even in these regimes, the envelope-based method maintains a slight advantage, suggesting that the benefits of dimension reduction are preserved even as the influence of privacy noise diminishes.

Figure 2 presents box plots of MSE and interval width across all 50 simulations for each setting. These plots visually reinforce the trends observed in Table 1: our method consistently

yields tighter distributions of error and uncertainty, indicating improved efficiency in inference under privacy. Together, these results demonstrate the practical utility of combining dimension reduction via envelope regression with Bayesian inference on privatized sufficient statistics.

Overall, these simulations demonstrate that the proposed envelope regression method provides meaningful improvements over the baseline in both estimation accuracy and uncertainty quantification while preserving differential privacy.

5 Conclusion

In this work, we developed a differentially private Bayesian framework for envelope regression by extending the data augmentation MCMC approach proposed by Ju et al. (2022). While Ju et al. (2022) introduced a general Gibbs sampling scheme for private Bayesian inference and demonstrated it in standard linear and log-linear regression settings, our contribution adapts this approach to the structured setting of envelope regression. This specialization leverages the dimensionality reduction inherent in the envelope model to improve inference efficiency under privacy constraints. Simulation results suggest that incorporating the envelope structure leads to more accurate coefficient estimates and tighter credible intervals, especially when the data size is small or the privacy budget is tight. By recovering the original coefficient vector via $\boldsymbol{\beta} = \mathbf{B}_1\boldsymbol{\theta}$, our method enables fair comparison with the baseline model, which does not account for immaterial variation in predictors.

An important area for further investigation is the calibration of uncertainty under privacy. The credible intervals produced by our method are valid Bayesian summaries, but they are not calibrated to guarantee frequentist coverage in the presence of privatized data. Although we did not compute empirical coverage rates in this study, understanding how privacy noise affects the frequentist properties of Bayesian credible intervals remains an open question and a valuable direction for future work. Beyond that, future extensions could address high-dimensional predictors, robust model misspecification, or privacy-aware selection of the envelope dimension. While the current framework is focused on a specific class of regression models, the approach may generalize to other structured settings where targeted dimension reduction can improve utility under differential privacy.

Supplementary Material

A compressed folder containing the code used to generate the results in Section 4 and to implement our proposed methods is available online.

A Appendix

A.1 MCMC Gibbs Sampler

The detailed MCMC Gibbs sampling approach for the parameters $\boldsymbol{\nu}$ and latent variables $(\mathbf{x}, \mathbf{y})$ mentioned in the model is given below.

Updating $\boldsymbol{\theta}$ With $\boldsymbol{\theta} \sim \mathcal{N}(0, 100\mathbf{I}_r)$, we have

$$\begin{aligned} p(\boldsymbol{\theta}|\sigma, \mathbf{x}, \mathbf{y}, \mathbf{B}_1) &\propto f(\mathbf{y}|\mathbf{x}, \mathbf{B}_1, \sigma, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\ &\propto \exp\left\{ -\frac{\sum_{i=1}^n (y_i - \boldsymbol{x}_i^\top \mathbf{B}_1 \boldsymbol{\theta})^2}{2\sigma^2} - \frac{\boldsymbol{\theta}^\top \boldsymbol{\theta}}{200} \right\}. \end{aligned}$$

Thus $[p(\boldsymbol{\theta}^{(t+1)}|\sigma^{(t)}, \mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \mathbf{B}_1^{(t)}] \sim \mathcal{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$ where

$$\begin{aligned} \boldsymbol{\mu}_\theta &= \boldsymbol{\Sigma}_\theta \mathbf{B}_1^{(t)} \mathbf{x}^{(t+1)^\top} \boldsymbol{y}^{(t+1)}, \\ \boldsymbol{\Sigma}_\theta &= \frac{1}{\sigma^{(t)2}} \mathbf{B}_1^{(t)} \mathbf{x}^{(t+1)^\top} \mathbf{x}^{(t+1)} \mathbf{B}_1^{(t)^\top} + \frac{1}{100} \mathbf{I}_r. \end{aligned}$$

Updating $\boldsymbol{\Omega}$ With $p(\boldsymbol{\Omega}) \sim \text{IW}(\mathbf{S}, s)$, we have

$$\begin{aligned} p(\boldsymbol{\Omega}|\boldsymbol{\lambda}) &\propto p(\boldsymbol{\Omega}) f(\boldsymbol{\lambda}|\boldsymbol{\Omega}) \\ &\propto |\boldsymbol{\Omega}|^{-(s+r+1)/2} \exp\left\{ -\frac{1}{2} \text{tr}\big(S\boldsymbol{\Omega}^{-1}\big) \right\} \times |\boldsymbol{\Omega}|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^n \boldsymbol{\lambda}_i^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}_i \right\}. \end{aligned}$$

Thus we have $[p(\boldsymbol{\Omega}^{(t+1)}|\boldsymbol{\lambda}^{(t+1)}] \sim \text{IW}(\tilde{\mathbf{S}}, \tilde{s})$ where

$$\tilde{s} = s + n, \quad \tilde{\mathbf{S}} = \mathbf{S} + \boldsymbol{\lambda}^{(t+1)^\top} \boldsymbol{\lambda}^{(t+1)}.$$

Updating $\boldsymbol{\Omega}_0$ With $p(\boldsymbol{\Omega}_0) \sim \text{IW}(\mathbf{S}_0, s_0)$, we have

$$\begin{aligned} p(\boldsymbol{\Omega}_0|\boldsymbol{\xi}) &\propto p(\boldsymbol{\Omega}_0) f(\boldsymbol{\xi}|\boldsymbol{\Omega}_0) \\ &\propto |\boldsymbol{\Omega}_0|^{-(s_0+p-r+1)/2} \exp\left\{ -\frac{1}{2} \text{tr}\big(S\boldsymbol{\Omega}_0^{-1}\big) \right\} \times |\boldsymbol{\Omega}_0|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^n \boldsymbol{\xi}_i^\top \boldsymbol{\Omega}_0^{-1} \boldsymbol{\xi}_i \right\}. \end{aligned}$$

Thus $[p(\boldsymbol{\Omega}_0^{(t+1)}|\boldsymbol{\xi}^{(t+1)}] \sim \text{IW}(\tilde{\mathbf{S}}_0, \tilde{s}_0)$ where

$$\tilde{s}_0 = s_0 + n, \quad \tilde{\mathbf{S}}_0 = \mathbf{S}_0 + \boldsymbol{\xi}^{(t+1)^\top} \boldsymbol{\xi}^{(t+1)}.$$

Updating σ^2 With $p(\sigma^2) \sim \text{IG}(a_0, b_0)$, we have

$$\begin{aligned} p\big(\sigma^2|\mathbf{x}, \mathbf{y}, \mathbf{B}_1, \boldsymbol{\theta}\big) &\propto p\big(\sigma^2\big) f(\mathbf{y}|\mathbf{x}, \mathbf{B}_1, \boldsymbol{\theta}) \\ &\propto \big(\sigma^2\big)^{-(a_0+1)} \exp\left\{ -\frac{b_0}{\sigma^2} \right\} \times \big(\sigma^2\big)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \big(y_i - \boldsymbol{x}_i^\top \mathbf{B}_1 \boldsymbol{\theta}\big)^2 \right\}. \end{aligned}$$

Thus, $[p((\sigma^2)^{(t+1)}|\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)}, \mathbf{B}_1^{(t)}, \boldsymbol{\theta}^{(t+1)}] \sim \text{IG}(\tilde{a}_0, \tilde{b}_0)$

$$\begin{aligned} \tilde{a}_0 &= a_0 + \frac{n}{2}, \\ \tilde{b}_0 &= b_0 + \frac{1}{2} \sum_{i=1}^n \big(y_i^{(t+1)} - \boldsymbol{x}_i^{(t+1)^\top} \mathbf{B}_1^{(t)} \boldsymbol{\theta}^{(t+1)}\big)^2. \end{aligned}$$

Updating $\mathbf{B}_1$ and $\mathbf{B}_2$ through $\mathbf{A}$ With $\mathbf{B}_1^{(t+1)}$ and $\mathbf{B}_2^{(t+1)}$ expressed as functions of a matrix $\mathbf{A}$ and $p(\mathbf{A}) \sim \mathcal{MN}(\boldsymbol{A}_0, 10\mathbf{I}, 10\mathbf{I})$, we have

$$
\begin{aligned}
&p(\mathbf{A}|\mathbf{x}, \mathbf{y}, \sigma, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\theta}) \\
&\quad \propto f\big(\mathbf{y}|\mathbf{x}, \mathbf{B}_1(\mathbf{A}), \sigma, \boldsymbol{\theta}\big) f\big(\mathbf{x}|\mathbf{B}_1(\mathbf{A}), \mathbf{B}_2(\mathbf{A}), \boldsymbol{\Omega}, \boldsymbol{\Omega}_0\big) p(\mathbf{A}) \\
&\quad \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(y_i - \boldsymbol{x}_i^\top \mathbf{B}_1\boldsymbol{\theta}\big)^2\right\} |\boldsymbol{\Sigma}_\mathbf{x}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\boldsymbol{x}_i^\top \boldsymbol{\Sigma}_x \boldsymbol{x}_i\right\} p(\mathbf{A}).
\end{aligned}
$$

Since there is no closed-form posterior, $\mathbf{A}$ is updated through the following Metropolis-Hastings algorithm.

1. Propose $\mathbf{A}^\star$, $\mathbf{A}^\star = \mathbf{A} + \boldsymbol{e}$ with $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\tau}^2)$.
2. Calculate $p_\mathbf{A} = p(\mathbf{A}|\mathbf{x}, \mathbf{y}, \sigma, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\theta})$ and $p_\mathbf{A}^\star = p(\mathbf{A}^\star|\mathbf{x}, \mathbf{y}, \sigma, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\theta})$.
3. Accept the proposed state with probability $\alpha(\mathbf{A}^\star) = \min\{\frac{p_\mathbf{A}^\star}{p_\mathbf{A}}, 1\}$.

Note that $\boldsymbol{\tau}$ serves as a hyper-parameter that is chosen to ensure that the acceptance rate of $\mathbf{A}$ falls within the specified interval of $(0.2, 0.6)$.

Use of Random Seeds To maintain traceable outcomes, we employed distinct seeds for each simulated dataset. For initializing the MCMC process, random values were utilized for the parameters $\boldsymbol{\nu}$ as well as the latent $(\mathbf{x}, \mathbf{y})$.

A.2 Statement on Computing Resources

We ran the experiments through software R on Compute Canada high performance cluster. We conducted individual MCMC chains, each comprising 10000 iterations. A standard chain necessitates approximately 9 minutes to complete when considering a sample size of $n = 500$, and around 30 minutes when $n = 1000$. The runtime for the largest sample size, $n = 5000$, increases substantially to around 12 hours. This increase is mainly due to increased rejection rates in the data augmentation step and higher computational cost associated with large matrix operations, such as inversion and multiplication, in each Gibbs sampling iteration. For smaller and moderate sample sizes, however, the computational burden remains manageable.

Funding

The research received funding from the Canada CIFAR AI Chairs program, the Alberta Machine Intelligence Institute, the Natural Sciences and Engineering Council of Canada, and the Canadian Statistical Sciences Institute.

References

Aoshima M, Shen D, Shen H, Yata K, Zhou YH, Marron JS (2018). A survey of high dimension low sample size asymptotics. *Australian & New Zealand Journal of Statistics*, 60: 4–19. https://doi.org/10.1111/anzs.12212

Aoshima M, Yata K (2017). Statistical inference for high-dimension, low-sample-size data. *American Mathematical Society, Sugaku Expositions*, 30: 137–158. https://doi.org/10.1090/suga/421

Balle B, Wang YX (2018). Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In: *International Conference on Machine Learning*, 394–403. PMLR.

Bernstein G, Sheldon D (2018). Differentially private Bayesian inference for exponential families. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2924–2934. Curran Associates Inc., Red Hook, NY, USA.

Bernstein G, Sheldon D (2019). Differentially private bayesian linear regression. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 525–535. Curran Associates Inc., Red Hook, NY, USA.

Chanyaswad T, Dytso A, Poor HV, Mittal P (2018). MVG mechanism: Differential privacy under matrix-valued query. In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 230–246. Association for Computing Machinery, New York, NY, USA.

Chaudhuri K, Sarwate AD, Sinha K (2013). A near-optimal algorithm for differentially-private principal components. *Journal of Machine Learning Research*, 14(1): 2905–2943.

Cook RD, Li B, Chiaromonte F (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, 20(3): 927–1010.

Cook RD, Zhang X (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association*, 110(510): 599–611. https://doi.org/10.1080/01621459.2014.983235

Dandekar A, Basu D, Bressan S (2018). Differential privacy for regularised linear regression. In: *International Conference on Database and Expert Systems Applications*, 483–491. Springer.

Doe J, Roe J (2021). Differential privacy techniques for census data analysis. *Journal of Census and Demographic Analysis*, 15(2): 123–137.

Dwork C, Kenthapadi K, McSherry F, Mironov I, Naor M (2006a). Our data, ourselves: Privacy via distributed noise generation. In: *Advances in Cryptology-EUROCRYPT 2006: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques. Proceedings* 25. St. Petersburg, Russia, May 28–June 1, 2006, 486–503. Springer.

Dwork C, McSherry F, Nissim K, Smith A (2006b). Calibrating noise to sensitivity in private data analysis. In: *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006. Proceedings 3*. New York, NY, USA, March 4–7, 2006, 265–284. Springer.

Dwork C, Roth A (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4): 211–407.

Dyda A, Purcell M, Curtis S, Field E, Pillai P, Ricardo K, et al. (2021). Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality. *Patterns*, 2(12). https://doi.org/10.1016/j.patter.2021.100366

Frühwirth-Schnatter S (2006). *Finite Mixture and Markov Switching Models*. Springer.

Ju N, Awan J, Gong R, Rao V (2022). Data augmentation MCMC for Bayesian inference from privatized data. *Advances in Neural Information Processing Systems*, 35: 12732–12743.

McSherry F, Mironov I (2009). Differentially private recommender systems: Building privacy into the Netflix prize contenders. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 627–636. Association for Computing Machinery, New York, NY, USA.

Smith A (2008). Efficient, differentially private point estimators. arXiv preprint: https://arxiv.org/abs/0809.4794.

Talwar K, Thakurta A, Zhang L (2015). Nearly-optimal private lasso. In: *Proceedings of the 29th International Conference on Neural Information Processing Systems*, 3025–3033. MIT Press, Cambridge, MA, USA.

Tierney L (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4): 1701–1728.

Wang D, Xu J (2019). On sparse linear regression in the local differential privacy model. In: *International Conference on Machine Learning*, 6628–6637. PMLR.

Yao Y, Li Z (2018). Differential privacy with bias-control limited sources. *IEEE Transactions on Information Forensics and Security*, 13(5): 1230–1241. https://doi.org/10.1109/TIFS.2017.2780802

Zhang Z, Rubinstein BIP, Dimitrakakis C (2016). On the differential privacy of bayesian inference. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2365–2371. AAAI Press.