Locally Private Causal Inference for Randomized Experiments

Yuki Ohnishi
Department of Statistics, Purdue University
and
Jordan Awan *
Department of Statistics, Purdue University

October 17, 2023

Abstract

Local differential privacy is a differential privacy paradigm in which individuals first apply a privacy mechanism to their data (often by adding noise) before transmitting the result to a curator. The noise for privacy results in additional bias and variance in their analyses. Thus it is of great importance for analysts to incorporate the privacy noise into valid inference. In this article, we develop methodologies to infer causal effects from locally privatized data under randomized experiments. First, we present frequentist estimators under various privacy scenarios with their variance estimators and plug-in confidence intervals. We show a naïve debiased estimator results in inferior mean-squared error (MSE) compared to minimax lower bounds. In contrast, we show that using a customized privacy mechanism, we can match the lower bound, giving minimax optimal inference. We also develop a Bayesian nonparametric methodology along with a blocked Gibbs sampling algorithm, which can be applied to any of our proposed privacy mechanisms, and which performs especially well in terms of MSE for tight privacy budgets. Finally, we present simulation studies to evaluate the performance of our proposed frequentist and Bayesian methodologies for various privacy budgets, resulting in useful suggestions for performing causal inference for privatized data.

Keywords: Rubin Causal Model, Local Differential Privacy, Debiased Estimators, Dirichlet Process Mixture

*This work was supported in part by the National Science Foundation (NSF) grants SES 2150615.

1 Introduction

Causal inference is a fundamental consideration across a wide range of domains in science, technology, engineering, and medicine. Researchers study experimental or observational data to unveil the causal effects of treatment assignment in an unbiased manner with valid uncertainty quantification. A traditional gold standard for performing causal inference is the classical randomized experiment (Imbens and Rubin, 2015). In this type of experiment, a great deal of control and precautions can be taken so as to eliminate events that would introduce instabilities and biases in causal inferences.

On the other hand, differential privacy (DP), introduced by Dwork et al. (2006), is another growing domain in science and business, as privacy protection has become a core concern for many organizations in the modern data-rich world. DP is a mathematical framework that provides a probabilistic guarantee that protects private information about individuals when publishing statistics about a dataset. This probabilistic guarantee is often achieved by adding random noise to the data. One DP model is the *central* differential privacy model, in which the data curators have access to the sensitive data and apply a DP mechanism to the data to produce the published outputs. A weakness of this model is that users are required to trust the data curators with their sensitive data. Another DP model is *local* differential privacy (LDP). In this model, the users do not directly provide their data to the data curator; instead, users apply the DP mechanism to their data locally before sending it to the curator. LDP is a preferable model if the data curators are not trusted by users. The LDP model has been adopted by various tasks and organizations, e.g., Google (Erlingsson et al., 2014) and Apple (Apple, 2017), for more stringent privacy protection.

Drawing causal conclusions from privatized data can be challenging. While the added random noise helps in safeguarding individuals' privacy, it distorts the actual patterns in the data. This distortion can lead to biased conclusions even in randomized experiments. This issue becomes even more pronounced in the LDP method, where each data point is individually altered before it is compiled. Therefore, when trying to understand cause-and-effect relationships using this protected data, researchers must exercise extra caution to ensure their interpretations remain accurate and unbiased.

In this article, we propose statistically valid causal inferential methodologies under three distinct local privacy scenarios. The first scenario, which we refer to as a "joint scenario", assumes that all accessible variables are separately privatized. In the second and third scenarios, which we term as "custom scenarios", we are allowed to select the variables we privatize with known and unknown treatment assignment probabilities. We then offer causal inference methodologies to analyze such privatized data. Our main contributions are as follows:

- We propose a "naïve" inverse probability weighting (IPW) estimator under the joint scenario. We compute the bias of the IPW estimator and propose a debiasing technique.
- We propose efficient frequentist estimators that achieve the minimax optimal rate under custom scenarios where we are allowed to select the variables we privatize.
- We also compute the asymptotic variance and construct asymptotic plugin nominal confidence intervals for all frequentist estimators. We discuss their optimality under each scenario.
- We develop a flexible and efficient Bayesian nonparametric methodology, along with a data augmentation Gibbs sampler tailored for locally privatized observations, which

can be applied to all scenarios that are considered in the frequentist analyses.

- We present simulation studies and empirical data analysis to evaluate the frequentist and Bayesian methodologies at various privacy budgets, resulting in useful suggestions for performing causal inference for privatized data.
- We propose a regression adjustment technique under the joint scenario in the Supplementary Materials. We show both theoretically and empirically that it helps improve the accuracy, but the gain is somewhat limited when the privacy budgets are tight.

The rest of the paper is organized as follows. Section 2 presents the preliminaries for the Rubin Causal Model and LDP. In Section 3, we develop frequentist approaches to inferring the causal effects of interest. Section 4 presents a Bayesian methodology for performing valid causal inference with the privatized data. Section 5 provides simulation studies for validating our methodologies developed in the previous sections, and Section 6 provides an application of our methodologies to real-world data of a cash transfer program conducted in Columbia. Section 7 concludes with some final discussion. The Supplementary Materials contains proofs, technical details, and additional numerical results.

1.1 Related Work

While DP is a rapidly growing field, the literature on causal inference methodologies for differentially privatized data remains sparse. The following work uses LDP for its DP mechanism. Agarwal and Singh (2021) introduced an end-to-end procedure for covariates cleaning, estimation, and inference, offering covariates cleaning-adjusted confidence intervals under the local differential privacy mechanism.

Some researchers have developed causal inference methodologies under the central DP model. D'Orazio et al. (2015) introduced the construction of central differential privacy mechanisms for summary statistics in causal inference. They then presented new algorithms for releasing differentially private estimates of causal effects and the generation of differentially private covariance matrices from which any least squares regression may be estimated. Lee et al. (2019) proposed a privacy-preserving inverse propensity score estimator for estimating the average treatment effect (ATE). Komarova and Nekipelov (2020) studied the impact of differential privacy on the identification of statistical models and demonstrated identification of causal parameters failed in regression discontinuity design under the central differential privacy. Niu et al. (2022) introduced a general meta-algorithm for privately estimating conditional average treatment effects. Kusner et al. (2016) tackles causal inference using a framework called the additive noise model (ANM), a more restrictive causal model than the Rubin Causal Model.

In non-causal domains, Evans and King (2022) offered statistically valid linear regression estimates and descriptive statistics for locally private data that can be interpreted as ordinary analyses of non-confidential data but with appropriately larger standard errors. Schein et al. (2019) presented an MCMC algorithm that approximates the posterior distribution over the latent variables conditioned on data that has been locally privatized by the geometric mechanism. Ju et al. (2022) proposed a general privacy-aware data augmentation MCMC framework to perform Bayesian inference from privatized data.

2 Preliminaries

2.1 Rubin Causal Model

Causal inference is of fundamental importance across many scientific and engineering domains that require informed decision-making based on experiments. Throughout this manuscript, we adopt the Rubin Causal Model (RCM) as our causal paradigm. In the RCM it is critical to first carefully define the Science of a particular problem, i.e., to define the experimental units, covariates, treatments, and potential outcomes (Imbens and Rubin, 2015). We consider N experimental units, indexed by $i = 1, \ldots, N$, that correspond to physical objects at a particular point in time. Each unit i has an observed outcome Y_i and treatment assignment W_i respectively. We consider a binary treatment $W_i \in \{0, 1\}$ with a fixed assignment probability, $p = P(W_i = 1)$, which is assumed to be known by the experimental design, and let $Y_i(w)$ denote a potential outcome for $w \in \{0, 1\}$. In this article, we consider the N units as a random sample from a large super-population, and we are interested in inferring the Population Average Treatment Effect (PATE): $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$. We invoke the common set of assumptions, which enable us to identify the PATE by the estimators derived in this manuscript (Imbens and Rubin, 2015).

Assumption 1. *1. (Positivity) The probability of treatment assignment given the covariates is bounded away from zero and one:* $0 < P(W_i = 1) < 1$.

2. (Random Assignment) The potential outcomes are independent of treatment assignment: $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i$.

3. (Stable Unit Treatment Value Assumption [SUTVA]) There is neither interference nor hidden versions of treatment. The observed outcome is formally expressed as: $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$.

2.2 Differential Privacy

In this article, we use the local differential privacy (LDP) model. Let $\mathcal{D}$ be the set of possible contributions from one individual in database D. In this paper, we only consider non-interactive local DP mechanisms. LDP is formally defined for any $\mathcal{D}$ as follows.

Definition 1 (Local Differential Privacy)**.** *An algorithm $\mathcal{M}$ is said to be ϵ-locally differentially private (ϵ-LDP) if for any two data points $x, x' \in \mathcal{D}$, and any $S \subseteq \mathrm{Range}(\mathcal{M})$,*

$$P(\mathcal{M}(x) \in S) \leq \exp(\epsilon) P(\mathcal{M}(x') \in S).$$

Intuitively, if an individual were to change their value from x to x', the output distribution of M would be similar, making it difficult for an adversary to determine whether x or x' was the true value. The value ϵ is called the *privacy budget* and lower values indicate a stronger privacy guarantee. Two important properties of differential privacy are *composition* and *invariance to post-processing*. Composition allows one to derive the cumulative privacy cost when releasing the results of multiple privacy mechanisms: if $\mathcal{M}_1$ is ϵ_1-LDP and $\mathcal{M}_2$ is ϵ_2-DP, then the joint release $(\mathcal{M}_1(x), \mathcal{M}_2(x))$ satisfies $(\epsilon_1 + \epsilon_2)$-LDP. Invariance to post-processing ensures that applying a data-independent procedure to the output of a DP mechanism does not compromise the privacy guarantee: if $\mathcal{M}$ is ϵ-LDP with range $\mathcal{Y}$, and $f : \mathcal{Y} \to \mathcal{Z}$ is a (potentially randomized) function, then $f \circ \mathcal{M}$ is also ϵ-LDP. Invariance to post-processing is especially important in this paper, as all of our inference procedures can be expressed as post-processing of more basic DP quantities.

One of the most commonly used DP mechanisms is the Laplace mechanism, which adds noise to a function of interest. Importantly, the noise must be scaled proportionally to the *sensitivity* of the function, which measures the worst-case magnitude by which the function's value may change between two individuals. Formally, the ℓ_1-sensitivity of a function f: $\mathcal{D} \to \mathbb{R}^k$ is $\Delta_f = \sup_{x,y \in \mathcal{D}} ||f(x) - f(y)||_1$.

Proposition 1 (Laplace Mechanism). *Let $f : \mathcal{D} \to \mathbb{R}^k$. The Laplace mechanism is defined as $M(x) = f(x) + (\nu_1, ..., \nu_k)^\top$, where the ν_i are independent Laplace random variables, $\nu_i \sim \text{Lap}(0, \Delta f/\epsilon)$, where the density of the Laplace distribution, $\text{Lap}(\mu, b)$, is given by $f(\nu|\mu, b) = \frac{1}{2b}\exp(-\frac{|\nu-\mu|}{b})$. Then M satisfies ϵ-LDP.*

For a binary variable (e.g., treatment assignment), a common mechanism is the randomized response.

Proposition 2 (Randomized Response Mechanism). *Let $Z_i \in \{0, 1\}$ be a binary variable. The randomized response mechanism is defined as*

$$M(Z_i) = \begin{cases} Z_i & w.p.\ \frac{\exp(\epsilon)}{1+\exp(\epsilon)} \\ 1 - Z_i & w.p.\ \frac{1}{1+\exp(\epsilon)}, \end{cases}$$

which satisfies ϵ-LDP.

3 Frequentist Approach

3.1 Minimax Risk Lower Bound for PATE Estimation

In this section, we discuss frequentist estimators for τ under several privacy scenarios where variables are privatized in different manners. According to Duchi et al. (2018), the minimax lower bound of the mean-squared error (MSE) for one-dimensional mean estimation is $O((N\epsilon^2)^{-1})$. In Lemma 1, we show that this same lower bound applies to the MSE for PATE estimation as well. We let $\mathcal{M}_\epsilon$ denote the set of all privacy mechanisms that satisfy ϵ-LDP. To ensure bounded ℓ_1-sensitivity, we assume $Y_i(w) \in [0, 1]$ for $i = 1, \ldots, N$, and $\{Y_i(w)\}_{i=1}^N$ are drawn according to some distribution $P_w \in \mathcal{P}_w$, where $\mathcal{P}_w$ denotes a class of distributions on the sample space of potential outcomes. Our restriction to $Y_i(w) \in [0, 1]$ is for simplicity and clarity. This follows standard practice (e.g., Lei et al. (2017), Ferrando et al. (2022) to name a few), and our discussions can be easily generalized to bounded outcomes $Y_i(w) \in [a, b]$ with $-\infty < a < b < \infty$ using shifting and scaling factors. We define an estimator $\hat{\tau}$ as a measurable function that maps privatized inputs to a real value, that is, $\hat{\tau} : \mathcal{X}^N \to \mathbb{R}$, where $\mathcal{X}$ generally denotes the space of privatized inputs under various privacy scenarios.

Lemma 1. *For $\epsilon \in [0, 1]$, there exists a constant c such that*

$$c \min(1, (N\epsilon^2)^{-1}) \leq \inf_{M_\epsilon \in \mathcal{M}_\epsilon} \inf_{\hat{\tau}} \sup_{\substack{P_0 \in \mathcal{P}_0, \\ P_1 \in \mathcal{P}_1, \\ p \in [0,1]}} \mathbb{E}[(\hat{\tau} - \tau)^2] \tag{1}$$

Lemma 1 implies that the optimal estimator of the PATE estimation problem also has the minimax lower bound $O((N\epsilon^2)^{-1})$.

3.2 Joint Scenario with Known p

We first consider a scenario where all variables are jointly and separately privatized. The observed outcomes are privatized by the Laplace mechanism. The privatized outcomes are $\tilde{Y}_i = Y_i + \nu_i^Y$, where $\nu_i^Y \sim \text{Lap}(1/\epsilon_y)$. The binary treatment variable W_i is privatized by the random response mechanism.

$$\tilde{W}_i = \begin{cases} W_i & \text{w.p. } q_{\epsilon_w} = \frac{\exp(\epsilon_w)}{1+\exp(\epsilon_w)} \\ 1 - W_i & \text{w.p. } 1 - q_{\epsilon_w} = \frac{1}{1+\exp(\epsilon_w)}. \end{cases}$$

By composition, the joint release of $(\tilde{Y}_i, \tilde{W}_i)_{i=1}^N$ satisfies $(\epsilon_y + \epsilon_w)$-LDP. $\tilde{Y}_i$ is observed after adding noise to Y_i, which is either $Y_i(0)$ or $Y_i(1)$, but we cannot identify which it is through the observed variables because W_i is also unobserved.

First, we propose estimators by plugging in the privatized observations into classical formulas, then derive bias correction results of the plug-in estimators. We also provide variance estimators, enabling asymptotically accurate plug-in confidence intervals.

We consider the following naïve inverse probability weighting (IPW) estimator $\tilde{\tau}_{naive}$. This naïve IPW estimator is defined by plugging in privatized observations for the usual IPW estimator.

$$\tilde{\tau}_{naive} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\tilde{W}_i\tilde{Y}_i}{\rho_1} - \frac{(1-\tilde{W}_i)\tilde{Y}_i}{\rho_0}\right\}, \tag{2}$$

where $\rho_w = P(\tilde{W}_i = w)$ for $w = 0, 1$. Note that ρ_w is a known marginal probability expressed by p and q_{ϵ_w}. The following lemma quantifies the bias of the estimator (2).

Lemma 2. *Under Assumption 1, the estimator (2) is biased for τ. The bias is*

$$\text{Bias}(\tilde{\tau}_{naive}) = \left(\frac{1}{C_{p,\epsilon_w}} - 1\right)\tau,$$

where $C_{p,\epsilon_w} = \frac{\rho_0\rho_1}{p(1-p)(2q_{\epsilon_w}-1)}$ with $q_{\epsilon_w} = \exp(\epsilon_w)/(1+\exp(\epsilon_w))$.

Let $\hat{E}_w = \frac{1}{\tilde{N}_w}\sum_{i:\tilde{W}_i=w}\tilde{Y}_i$ and $\hat{V}_w = \frac{1}{\tilde{N}_w-1}\sum_{i:\tilde{W}_i=w}(\tilde{Y}_i - \hat{E}_w)^2$, where $\tilde{N}_w = \sum_{i=1}^N \mathbb{1}(\tilde{W}_i = w)$ for $w = 0, 1$. In Theorem 3.1, we show that the estimator $C_{p,\epsilon_w}\tilde{\tau}_{naive}$ is unbiased, consistent, and that we can construct asymptotically valid confidence intervals for PATE based on this estimator.

Theorem 3.1. 1. *(Unbiasedness & Consistency) $C_{p,\epsilon_w}\tilde{\tau}_{naive}$ is unbiased and consistent for τ.*

2. *(CLT) $\sqrt{N}(C_{p,\epsilon_w}\tilde{\tau}_{naive} - \tau)$ converges in distribution to a mean-zero normal distribution.*
3. *(Confidence Interval) The following interval is the nominal central confidence at the*

significance level α:

$$\left(C_{p,\epsilon_w}\tilde{\tau}_{naive} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{naive}}{N}}, C_{p,\epsilon_w}\tilde{\tau}_{naive} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{naive}}{N}}\right),$$

where $\hat{\Sigma}_{naive} = C^2_{p,\epsilon_w}(\frac{1}{\rho_1}\hat{V}_1 + \frac{1}{\rho_0}\hat{V}_0 + \frac{\rho_0}{\rho_1}\hat{E}_1^2 + \frac{\rho_1}{\rho_0}\hat{E}_0^2 + 2\hat{E}_0\hat{E}_1)$.

4. *(Convergence rate) The MSE of $C_{p,\epsilon_w}\tilde{\tau}_{naive}$ is $O((N\epsilon_y^2\epsilon_w^2)^{-1})$.*

The details of the asymptotic normality and the confidence interval construction are in Supplementary Material A.4. Setting $\epsilon_y = \epsilon_w = \epsilon/2$ gives MSE of $O((N\epsilon^4)^{-1})$, which matches the minimax rate (1) in terms of N, but not in terms of ϵ. In the following sections, we see that when we use a customized privacy mechanism, rather than a naïve joint privatization, we can match the minimax lower bound. In the Supplementary Materials, we introduce another class of frequentist estimators: the OLS estimator, specifically under the joint scenario. We explore both the advantages and limitations of the OLS estimator in comparison to the IPW estimator within this context.

3.3 Custom Scenario with Known p

In this section, we will tailor the privacy mechanism to the PATE estimation problem, assuming that the value p is known (such as in most designed experiments). Specifically, for unit $i = 1, \ldots, N$, we privatize the following variable by the Laplace mechanism: $A_i = \frac{W_iY_i}{p} - \frac{(1-W_i)Y_i}{1-p}$. The sensitivity of A is $\Delta_A = \max(\frac{1}{p}, \frac{1}{1-p})$. The privatized value of A is $\tilde{A}_i = A_i + \nu_i^A$, where $\nu_i^A \sim \text{Lap}(\Delta_A/\epsilon_a)$. Then, it is straightforward to show that the following IPW estimator is unbiased and consistent.

$$\tilde{\tau}_{IPW} = \frac{1}{N}\sum_{i=1}^{N}\tilde{A}_i. \tag{3}$$

Theorem 3.2. *1. (Unbiasedness & Consistency) $\tilde{\tau}_{IPW}$ is unbiased and consistent for τ.*

2. *(CLT) $\sqrt{N}(\tilde{\tau}_{IPW} - \tau)$ converges in distribution to a mean-zero normal distribution.*
3. *(Confidence Interval) The following interval is the nominal central confidence at the significance level α:*

$$\left(\tilde{\tau}_{IPW} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{IPW}}{N}}, \tilde{\tau}_{IPW} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{IPW}}{N}}\right),$$

where $\hat{\Sigma}_{IPW} = \frac{1}{N-1}\sum_{i=1}^{N}(\tilde{A}_i - \hat{E}_A)^2$ with $\hat{E}_A = \frac{1}{N}\sum_{i=1}^{N}\tilde{A}_i$.

4. *(Convergence rate) The MSE of $\tilde{\tau}_{IPW}$ is $O((N\epsilon_a^2)^{-1})$.*

The details of the asymptotic normal distribution and the confidence interval construction are provided in Supplementary Material A.5. We see in Theorem 3.2 that the lower bound of the IPW estimator under the custom scenario matches the minimax lower bound for the locally private PATE estimation (1), improving over the naïve estimator from Section 3.2.

3.4 Custom Scenario with Unknown p

The estimator (3) is appealing in the sense of optimality when p is known, such as in randomized experiments, however, their application is restricted when p is unknown. In this regard, we proceed a step further to address situations in which p is inaccessible, while Assumption 1 remains valid. Examples of this setting include A/B testing and clinical trials, where marketers or doctors assign treatments with an undisclosed probability (that does not depend on the covariate information).

We consider releasing the following quantities: $\tilde{\mathbf{B}}_i = (\tilde{B}_{i,1}, \tilde{B}_{i,2}, \tilde{B}_{i,3})$, where

$$\tilde{B}_{i,1} = W_i Y_i + \nu_i^{B_1}, \tilde{B}_{i,2} = (1 - W_i) Y_i + \nu_i^{B_2}, \text{ and } \tilde{B}_{i,3} = W_i + \nu_i^{B_3},$$

where $\nu_i^{B_j} \sim \text{Lap}(1/\epsilon_{b_j})$ for $j = 1, 2, 3$. We also let $\tilde{B}_{i,4} = 1 - \tilde{B}_{i,3}$. By composition, the joint release of $(\tilde{B}_{i,1}, \tilde{B}_{i,2}, \tilde{B}_{i,3})_{i=1}^N$ satisfies $(\epsilon_{b_1} + \epsilon_{b_2} + \epsilon_{b_3})$-LDP.

Given these privatized quantities, we construct our difference-in-means (DM) estimator as follows.

$$\tilde{\tau}_{DM} = \frac{\sum_{i=1}^N \tilde{B}_{i,1}}{\sum_{i=1}^N \tilde{B}_{i,3}} - \frac{\sum_{i=1}^N \tilde{B}_{i,2}}{\sum_{i=1}^N \tilde{B}_{i,4}}. \tag{4}$$

Let $\hat{E}_{B_j} = \frac{1}{N} \sum_{i=1}^N \tilde{B}_{i,j}$, $\hat{V}_{B_j} = \frac{1}{N-1} \sum_{i=1}^N (\tilde{B}_{i,j} - \hat{E}_{B_j})^2$ for $j = 1, 2, 3, 4$ and $\widehat{\text{Cov}}_{j,k} = \frac{1}{N-1} \sum_{i=1}^N (\tilde{B}_{i,j} - \hat{E}_{B_j})(\tilde{B}_{i,k} - \hat{E}_{B_k})$ for $j \neq k$. We have the following properties for $\tilde{\tau}_{DM}$:

Theorem 3.3. *1. (Consistency) $\tilde{\tau}_{DM}$ is consistent for τ.*

2. (CLT) $\sqrt{N}(\tilde{\tau}_{DM} - \tau)$ converges in distribution to a mean-zero normal distribution.

3. (Confidence Interval) The following interval is the nominal central confidence at the significance level α:

$$\left(\tilde{\tau}_{DM} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\Sigma}_{DM}}{N}}, \tilde{\tau}_{DM} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\Sigma}_{DM}}{N}} \right),$$

where $\hat{\Sigma}_{DM} = \hat{\mathbf{e}}' \hat{\mathbf{S}} \hat{\mathbf{e}}$, with $\hat{\mathbf{e}} = (1/\hat{E}_{B_3}, -1/(1 - \hat{E}_{B_3}), -\hat{E}_{B_1}/\hat{E}_{B_3}^2, \hat{E}_{B_2}/(1 - \hat{E}_{B_3})^2)'$ and

$$\hat{\mathbf{S}} = \begin{pmatrix} \hat{V}_{B_1} & \widehat{\text{Cov}}_{1,2} & \widehat{\text{Cov}}_{1,3} & \widehat{\text{Cov}}_{1,4} \\ \widehat{\text{Cov}}_{2,1} & \hat{V}_{B_2} & \widehat{\text{Cov}}_{2,3} & \widehat{\text{Cov}}_{2,4} \\ \widehat{\text{Cov}}_{3,1} & \widehat{\text{Cov}}_{3,2} & \hat{V}_{B_3} & \widehat{\text{Cov}}_{3,4} \\ \widehat{\text{Cov}}_{4,1} & \widehat{\text{Cov}}_{4,2} & \widehat{\text{Cov}}_{4,3} & \hat{V}_{B_4} \end{pmatrix}.$$

4. (Convergence rate) The MSE of $\tilde{\tau}_{DM}$ is $O((N(\epsilon_{b_1}^2 + \epsilon_{b_2}^2 + \epsilon_{b_3}^2))^{-1})$.

The details of the asymptotic normal distribution and the confidence interval construction are provided in Supplementary Material A.6. Setting $\epsilon_{b_1} = \epsilon_{b_2} = \epsilon_{b_3} = \epsilon/3$ gives $O((N\epsilon^2)^{-1})$, which also matches the minimax lower bound of (1), indicating the optimality of the estimator.

3.5 Discussion on Frequentist Estimators

The three scenarios serve different purposes. While the joint scenario permits the release of the entire synthetic dataset to analysts, it suffers from the privatization of multiple variables, thereby compromising its optimality. As discussed in the Supplementary Materials, the OLS estimator helps improve the efficiency under the joint scenario, however, the gain is limited since we must pay additional privacy budgets for covariates. In the custom scenarios, access to the complete dataset is unavailable, but the estimators attain the optimal rate of the locally private PATE estimation. While both custom estimators achieve the minimax rate, the estimator with known p is able to focus its privacy budget on a single quantity, which gives improved finite sample performance; see Section 5.

When the sample size is small, or when privacy budgets are too tight, it is possible that the point estimators and interval estimators are out of support of the estimand, as the estimand is assumed to be bounded, but the observed private data are usually unbounded. Therefore, we apply additional post-processing to clamp estimators to the closest end of the support when they are out of bounds. For example, if the initial estimator is $\hat{\tau} = 1.8$, then we instead set $\hat{\tau} = 1.0$. However, suppose the lower and upper bounds of the estimated confidence interval are both clamped to the bounds of the support: in this case, the estimated confidence interval is not useful at all. This is a limitation of frequentist estimators arising from the trade-off between privacy and the accuracy of the analysis. This clamping processing is not necessary to achieve all the statistical properties derived in the paper. It only serves to reduce the MSE of the estimator by projecting the out-of-bound estimator to the bound.

4 Bayesian Approach

4.1 Overview of the Bayesian Methodology

Following the Bayesian paradigm of Rubin (1978), we consider deriving the posterior distributions of the causal estimands (Forastiere et al., 2016; Ohnishi and Sabbaghi, 2022a). The key idea is the data augmentation (Tanner and Wong, 1987) to obtain the posterior distribution of the causal estimands by imputing in turn the missing variables. The idea for estimating causal effects in the Bayesian paradigm is outlined in Rubin (1978); Imbens and Rubin (2015), but our unique challenges lie in the fact that neither treatment variable W nor either potential outcome $Y(0), Y(1)$ is observed.

To show how Bayesian inference proceeds in our framework, consider the following joint distribution of all observed variables $\tilde{\mathbf{O}}$ and missing variables $\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W}$: $P(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W}, \tilde{\mathbf{O}})$, where $\tilde{\mathbf{O}} = (\tilde{\mathbf{Y}}, \tilde{\mathbf{W}})$ for the joint scenario and $\tilde{\mathbf{O}} = \tilde{\mathbf{A}}$ or $\tilde{\mathbf{B}}$ for the custom scenarios. As discussed in Section D, since causal effects are identifiable under randomization without covariate adjustment and incorporating covariates requires additional privacy costs for their release, we do not include covariates in our Bayesian methodologies, but the extension should be straightforward (e.g., Maceachern (1999)). In what follows, we focus on the joint scenario discussed in Section 3.2 to show the outline of our algorithm, but it can easily be extended to the custom scenarios, as explained in Supplementary Material.

Under the super-population perspective, the observed and missing variables are considered as a joint draw from the population distribution. Bayesian inference considers the observed values of these quantities to be realizations of random variables and the missing values to be unobserved random variables. We also assume these quantities are unit exchangeable, then de Finetti's theorem implies that there exists a vector of parameters, $\boldsymbol{\theta}$, with the prior distribution $P(\boldsymbol{\theta})$ such that

$$\begin{aligned} P(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{W}, \tilde{\mathbf{Y}}, \tilde{\mathbf{W}}) &= \int P(\boldsymbol{\theta}) \prod_i P(Y_i(0), Y_i(1), W_i, \tilde{Y}_i, \tilde{W}_i \mid \boldsymbol{\theta}) d\boldsymbol{\theta} \\ = \int P(\boldsymbol{\theta}) \prod_i P(W_i) P(\tilde{W}_i \mid W_i) &P(Y_i(0), Y_i(1) \mid \boldsymbol{\theta}) P(\tilde{Y}_i \mid Y_i(0), Y_i(1), W_i) d\boldsymbol{\theta}, \end{aligned} \tag{5}$$

which follows from the conditional independence of potential outcomes and $\tilde{W}_i$ given W_i (Lemma 3 in the Supplementary Materials) and the random assignment assumption. The distribution of $\tilde{Y}_i$ depends not only on $Y_i(0)$ and $Y_i(1)$ but also on W_i because the DP mechanism is applied to the observed outcome $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$. Note that we know the DP mechanisms for W and Y, that is, $P(\tilde{Y}_i \mid Y_i(0), Y_i(1), W_i)$ and $P(\tilde{W}_i \mid W_i)$ have a known functional form. Therefore, the modeling effort is only required for $P(Y_i(0), Y_i(1) \mid \boldsymbol{\theta})$. Under this modeling strategy, our Bayesian approach is a valid inference for PATE. Note that PATE is a function of the parameters $\boldsymbol{\theta}$, which governs the potential outcomes. Thus, it suffices to obtain the posterior draws of the posterior of the $\boldsymbol{\theta}$ for the posterior draws of PATE.

A significant insight from (5) is that the treatment assignment mechanism is *not* ignorable. In conventional non-private settings, the treatment assignment model, represented as $P(W_i)$, is ignorable and falls out of the likelihood in Bayesian causal inference under randomization or unconfoundedness assumptions (Li et al., 2023). Yet, in a DP context, these treatment assignment variables are not directly observed. This necessitates the integration of both the treatment assignment models and their respective privacy mechanisms into our inferences. Additionally, a nuanced but crucial point is the necessity to model both $Y_i(0)$ and $Y_i(1)$. Typically, Bayesian causal inference for PATE estimation is performed via observable data (e.g., Zigler (2016); Stephens et al. (2023)). This is because the missing potential outcome eventually gets marginalized out from (5) under the assumption of prior parameter independence and unconfounded assignment, thus it does not influence parameter inference. In our scenario, however, it is uncertain whether $Y_i(0)$ or $Y_i(1)$ has been privatized to yield $\tilde{Y}_i$. This uncertainty calls for a data augmentation strategy for both potential outcomes.

We adopt the Dirichlet Process Mixture (DPM) to model $P(Y_i(0), Y_i(1) \mid W_i, \boldsymbol{\theta})$ for its flexibility. The DPM is a natural Bayesian choice for density estimation problems, which fits our needs that require $P(Y_i(0), Y_i(1) \mid W_i, \boldsymbol{\theta})$ to be estimated without assuming strong parametric forms. The following section and Supplementary Materials B provide technical details of the DPM and the Gibbs sampler.

4.2 Algorithm Outlines

Equation (5) motivates the Gibbs sampling procedures to obtain the draws from the posterior distribution of $\boldsymbol{\theta}$. This section describes the key steps of the Gibbs sampler. Each step is derived from the corresponding components of (5). For inference of DPM parameters, denoted by $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{u})$, we adopt an approximated blocked Gibbs sampler based on the truncation of the stick-breaking representation (Ishwaran and Zarepour, 2000), due to its simplicity. In this algorithm, we set a conservatively large upper bound, $K \leq \infty$, on the number of components that units potentially belong to. Let $C_i \in \{1, ..., K\}$ denote the latent class indicators with a multinomial distribution, $C_i \sim \text{Multinomial}(\mathbf{u})$ where $\mathbf{u} = (u_1, ..., u_K)$ denote the weights of all components of the DPM. More specific details about the DPM are provided in the Supplementary Material. The algorithm proceeds as follows.

1. Given $Y_i(0), Y_i(1)$, draw each W_i from $P(W_i = 1|-) = \frac{r_1}{r_0+r_1}$, where $r_w = P(\tilde{Y}_i \mid Y_i(w))P(\tilde{W}_i \mid W_i = w)P(W_i = w)$ for $w = 0, 1$.
2. Given $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{u}$, C_i and W_i, draw each $Y_i(0)$ and $Y_i(1)$ according to:

$$P(Y_i(W_i)|-) \propto P(Y_i(W_i) \mid \mu_{W_i}^{C_i}, \Sigma_{W_i}^{C_i})P(\tilde{Y}_i \mid Y_i(W_i))$$
$$P(Y_i(1-W_i)|-) \propto P(Y_i(1-W_i) \mid \mu_{1-W_i}^{C_i}, \Sigma_{1-W_i}^{C_i}).$$

3. Given $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{u}$, $Y_i(0)$ and $Y_i(1)$, draw each C_i from

$$P(C_i = k|-) \propto u_k P(Y_i(0) \mid \mu_0^k, \Sigma_0^k)P(Y_i(1) \mid \mu_1^k, \Sigma_1^k).$$

4. Let $u'_K = 1$. Given α, $\mathbf{C}$, draw u'_k for $k \in \{1, ..., K-1\}$ from

$$P(u'_k|-) \propto \text{Beta}\left(1 + \sum_{i:C_i=k} 1, \alpha + \sum_{i:C_i>k} 1\right).$$

   Then, update $u_k = u'_k \prod_{j<k}(1 - u'_j)$.
5. Given $\mathbf{C}$ and $\mathbf{u}'$, draw α from

$$P(\alpha|-) \propto P(\alpha) \prod_{k=1}^{K} f\left(u'_k \middle| 1 + \sum_{i:C_i=k} 1, \alpha + \sum_{i:C_i>k} 1\right),$$

   where f is the pdf of u'_k, the beta distribution. The standard Metropolis-Hastings algorithm is used for this step.
6. Given $\mathbf{Y}(0)$, $\mathbf{Y}(1)$ and $\mathbf{C}$, draw $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from

$$P(\mu_0^k, \Sigma_0^k|-) \propto H(\mu_0^k, \mu_1^k, \Sigma_0^k, \Sigma_1^k) \prod_{i:C_i=k} P(Y_i(0), Y_i(1) \mid \mu_0^k, \mu_1^k, \Sigma_0^k, \Sigma_1^k).$$

Remark. *The key steps of this algorithm are 1 and 2, which correspond to the data augmentation steps, imputing the latent variables $Y_i(0), Y_i(1)$ and W_i. In Step 1, the probability $P(\tilde{Y}_i \mid Y_i(w))$ for $w = 0, 1$ indicates that $\tilde{Y}_i$ is observed via privatizing the potential outcome $Y_i(w)$, which would have been observed if we observed $W_i = w$. In step 2, given W_i, the corresponding potential outcome $Y_i(W_i)$ is considered to be privatized, but the other missing potential outcome $Y_i(1 - W_i)$ should not be associated with the observed $\tilde{Y}_i$ within the iteration. Therefore, the posterior distribution of $Y_i(W_i)$ cannot be obtained in a closed form as it is weighted by the privacy mechanism $P(\tilde{Y}_i \mid Y_i(W_i))$, whereas the missing potential outcomes $Y_i(1 - W_i)$ are just generated from the outcome model $P(Y_i(1 - W_i) \mid \boldsymbol{\theta})$. We adopt the privacy-aware Metropolis-within-Gibbs algorithm proposed in Ju et al. (2022) for the posterior draws of $Y_i(W_i)$. They proposed a generic data augmentation approach of* *updating confidential data that exploits the privacy guarantee of the mechanism to ensure efficiency. Their algorithm has guarantees on mixing performance, indicating that the acceptance probability is lower bounded by* $\exp(-\epsilon_y)$. *Another advantage of their approach is that we may utilize the outcome model to sample a proposal value from* $P(Y_i(W_i)|\theta)$ *at the current value of* θ, *rather than specifying a custom proposal distribution and step size for the Metropolis-Hastings step. Finally, Steps 3–6 updates all the parameters of the DPM that govern the potential outcomes, using standard techniques; see Section B of the Supplementary Materials for details of the DPM, full details of the algorithm and the extension of the algorithm to the custom scenarios, which requires slight modifications to Steps 1 and 2.*

5 Simulation Studies

We evaluate the frequentist properties of our methodologies for various privacy budgets. The evaluation metrics that we consider are bias and mean square error (MSE) in estimating a causal estimand, coverage of an interval estimator for a causal estimand, and the interval length. Bias, MSE and coverage are generally defined as $\sum_{m=1}^{M} (\tau - \hat{\tau}_m)/M$, $\sum_{m=1}^{M} (\tau - \hat{\tau}_m)^2/M$ and $\sum_{m=1}^{M} \mathbb{1}\left(\hat{\tau}_m^l \leq \tau \leq \hat{\tau}_m^u\right)/M$ respectively, where M denotes the number of simulated datasets, τ denotes the true causal estimand, $\hat{\tau}_m$, $\hat{\tau}_m^l$ and $\hat{\tau}_m^u$ denote the estimate of the causal estimand, 95% lower and upper end of the interval estimator of the causal estimand using dataset $m = 1, \ldots, M$. Our summary of the interval length is the mean of the lengths of the intervals computed from M simulated datasets. For our Bayesian method, the point estimator is the mean of the posterior distribution of a causal estimand, and the interval estimator is the 95% central credible interval. We ran the MCMC algorithm for $100,000$ iterations using a burn-in of $50,000$. The iteration numbers were chosen after experimentation to deliver stable results over multiple runs.

5.1 Data-generating Mechanisms

For our simulations, we consider a Bernoulli randomized experiment with treatment assignment and covariates for unit i generated according to:

$$W_i \sim \text{Bernoulli}(0.5), X_{i,1} \sim \text{Uniform}(0,1), X_{i,2} \sim \text{Beta}(2,5), X_{i,3} \sim \text{Bernoulli}(0.7).$$

To generate potential outcomes, we adopt the Beta regression Ferrari and Cribari-Neto (2004): $Y_i(w) \sim \text{Beta}(\mu_i(w)\phi, (1-\mu_i(w))\phi)$, where $\mu_i(w)$ and ϕ are a location parameter and scale parameter respectively with $\mu_i(w) = \text{expit}(1.0 - 0.8X_1 + 0.5X_2 - 2.0X_3 + 0.5w)$ and $\phi = 50$. We consider $X_{i,d}$ to generate Y_i but do not release the privatized $\tilde{X}_{i,d}$. This model is beneficial for our simulations because the generated data automatically satisfy the following sensitivity: $\Delta_Y = 1$. Then, we obtain the private data $\tilde{Y}_i$, $\tilde{W}_i$, $\tilde{A}_i$, $\tilde{\mathbf{B}}_i$ by applying the corresponding privacy mechanisms. The actual value of PATE can be obtained in a closed form, which is necessary to calculate bias, MSE, and coverage. The details are provided in the Supplementary Materials.

Table 1: Evaluation metrics for IPW estimator under different privacy scenarios ($N = 10000, N_{sim} = 2000$). N_{sim} denotes the number of simulations. ϵ_{tot} denotes the total privacy budget. "Custom (IPW)" and "Custom (DM)" columns are scenarios in Section 3.3 and 3.4 respectively.

	Coverage			Bias			MSE			Interval Width		
ϵ_{tot}	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)
0.1	94.55%	94.95%	99.8%	0.9025	−0.1873	0.9025	0.9872	0.0803	0.7608	1.889	1.091	1.988
0.3	94.7%	94.1%	98.05%	0.9025	−0.0221	−0.4396	0.7875	0.0091	0.2518	1.882	0.371	1.655
1.0	94.65%	94.6%	95.6%	−0.2171	−0.0086	−0.1498	0.0568	0.0009	0.0201	0.915	0.117	0.553
3.0	95.3%	95.0%	95.3%	−0.033	−0.0078	0.0076	0.0011	0.0002	0.0022	0.13	0.052	0.182
10.0	94.9%	94.95%	94.4%	0.0	0.003	0.0012	0.0001	0.0001	0.0002	0.043	0.038	0.057

Table 2: Evaluation metrics of Bayesian estimators for $N = 10000, N_{sim} = 1000$.

	Coverage			Bias			MSE			Interval Width		
ϵ_{tot}	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)
0.1	96.4%	93.8%	96.3%	−0.0949	−0.0772	−0.0951	0.0099	0.0079	0.0099	0.34	0.319	0.341
0.3	96.9%	94.5%	94.6%	−0.0953	−0.036	−0.0897	0.0099	0.0034	0.0094	0.342	0.22	0.334
1.0	93.4%	93.8%	92.2%	−0.0691	−0.0069	−0.0511	0.0077	0.0006	0.0055	0.32	0.096	0.263
3.0	93.2%	92.2%	94.2%	−0.0081	−0.0063	−0.0098	0.0006	0.0002	0.001	0.093	0.045	0.117
10.0	95.0%	93.5%	92.3%	−0.0023	−0.0027	−0.0045	0.0	0.0	0.0001	0.026	0.022	0.036

5.2 Results

Table 1 presents the performance evaluation of our estimators under different scenarios for $N = 10000$ with various privacy budgets for ϵ_{tot}. We let $\epsilon_{tot} = \epsilon_a = \epsilon_y + \epsilon_w = \epsilon_{b_1} + \epsilon_{b_2} + \epsilon_{b_3}$, where $\epsilon_y = \epsilon_w$ and $\epsilon_{b1} = \epsilon_{b2} = \epsilon_{b3}$. All scenarios achieve about 95% coverage, except for the custom scenario (DM) of $\epsilon_{\text{tot}} = 0.1, .03$, which has some over-coverage. This may be because the estimator for the asymptotic variance has a non-negligible estimation error with the finite samples. The simulations in this section rely on the results of Section 3.2, 3.3, and 3.4 to build confidence intervals. The fact that the intervals have correct 95% coverage indicates that the estimators 1) are in fact asymptotically normal, 2) are asymptotically unbiased, and 3) have the stated asymptotic variance. For bias and MSE, we observe smaller bias and MSE for larger privacy budgets. The custom scenario (IPW) yields lower MSE than the joint scenario, which is also consistent with the discussion of the optimality in Section 3.2, 3.3, and 3.4, but the difference becomes negligible as ϵ_{tot} increases.

When we have a tight privacy budget of $\epsilon_{\text{tot}} = 0.1, 0.3$, the length of the confidence intervals of the joint scenario are nearly 2, which is almost non-informative about the estimand. With strict budget constraints and a small sample size, the analysis results may tell us little about the estimands, even though their consistency and confidence intervals are statistically valid. This is an inevitable trade-off between privacy protection and the accuracy of the results. Custom (IPW) has the best finite sample performance, offering informative intervals and small bias and MSE for all privacy budgets.

Table 2 compares our Bayesian methodology under the three scenarios. We see that the Bayes estimator yields well-calibrated coverage probabilities and smaller MSE and bias for most cases. The differences in MSE between frequentist estimators and Bayesian estimators become negligible as ϵ_{tot} gets large ($\epsilon_{\text{tot}} = 3.0, 10.0$). When the privacy budget is tight, the Bayesian methodology outperforms the frequentist approach in all metrics. Specifically, the interval length of the Bayes estimator for $\epsilon_{\text{tot}} = 0.1$ is around 0.35 for all scenarios, which is informative enough about the estimands. In the Supplementary Materials, we provide additional simulation studies for smaller sample sizes, as well as those for the OLS estimator under the joint scenario.

Table 3: Empirical analysis evaluating privatized cash transfer programs in Colombia. In the "Non-private" columns, "Freq" represents the standard IPW estimator, while "Bayes" represents the standard Dirichlet process mixture models for non-private data.

	Non-private						Private																	
							Joint						Custom (IPW)						Custom (DM)					
	Freq			Bayes			Freq			Bayes			Freq			Bayes			Freq			Bayes		
ϵ_{tot}	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
0.1	0.006	-0.042	0.054	0.005	0.001	0.008	1.0	-1.0	1.0	0.011	-0.178	0.145	0.019	-1.0	1.0	0.072	-0.135	0.244	1.0	-1.0	1.0	0.032	-0.137	0.193
0.3	0.006	-0.042	0.054	0.005	0.001	0.008	-1.0	-1.0	1.0	0.049	-0.082	0.190	0.010	-0.367	0.386	0.160	-0.038	0.389	-0.581	-1.0	1.0	0.049	-0.148	0.238
1.0	0.006	-0.042	0.054	0.005	0.001	0.008	-0.169	-0.898	0.559	0.041	-0.022	0.111	0.006	-0.118	0.131	0.073	-0.018	0.137	0.131	-0.546	0.809	0.054	-0.124	0.169
3.0	0.006	-0.042	0.054	0.005	0.001	0.008	0.008	-0.116	0.131	0.018	-0.007	0.044	0.005	-0.061	0.072	-0.002	-0.018	0.015	-0.038	-0.248	0.170	0.048	0.004	0.098
10.0	0.006	-0.042	0.054	0.005	0.001	0.008	0.006	-0.051	0.064	0.009	0.0	0.018	0.008	-0.048	0.064	-0.002	-0.009	0.006	-0.006	-0.066	0.054	0.015	0.002	0.027

6 Real Data Analysis

We applied our methodology to a real-world causal inference task. We analyzed a randomized experiment that examined the impact of a cash transfer program on students' attendance rates (Barrera-Osorio et al., 2011). Conducted at San Cristobal in Colombia, the study recruited households with one to five school children, randomly assigning children to either participate in the cash transfer program or not with probability $p = 0.628$. The number of recruited students is $N = 5240$. With known treatment assignment, we assessed the treatment effect of the program on the attendance rate of the students, with eligible students receiving cash subsidies if they attended school at least 80% of the time in a given month.

We utilized the privatization techniques as outlined in Section 3, setting ϵ_{tot} to values of 0.1, 0.3, 1.0, 3.0, and 10.0. Our methodologies were then benchmarked against non-private baseline methods, which offer target values for our private estimates. For the non-private frequentist baseline, we employed the standard IPW estimator.

Table 6 presents point mean estimators alongside the lower (2.5%) and upper (97.5%) bounds for interval estimators across each methodology. For the interval estimators, we used central confidence intervals for the frequentist approach and credible intervals for the Bayesian approach. Both frequentist and Bayesian non-private interval estimators highlighted a positive interval, indicating a significant effect. The point estimates showed a 0.6% increase in the frequentist non-private approach and a more modest 0.5% increase in the Bayesian approach. Given these results, our expectation for the private methodologies is, at best, to approximate the non-private values, since better inferences are unlikely with privatized data. Note that as the experimental data is fixed, the only randomness in this study is the privacy mechanisms.

The point estimates for both frequentist and Bayesian methodologies are similar to their non-private results when $\epsilon_{\text{tot}} \geq 3.0$. In particular, we observe that the Custom (IPW) scenario results in the narrowest confidence intervals. In the joint and custom (DM) scenarios, the frequentist estimators deviated more from the non-private one, showing larger intervals. The frequentist methodologies yield non-informative intervals when the privacy budget is tightest $\epsilon_{\text{tot}} = 0.1$. The Bayesian methodology demonstrated strong performance across all scenarios. These observations align with our simulation studies, further validating the efficacy of our methodologies.

7 Concluding Remarks

In this article we proposed causal inferential methodologies to analyze differential private data under the Rubin Causal Model. We considered three distinct local privacy scenarios that have practical relevance: 1) jointly privatized variables with known p, 2) custom privatized variables with known p, and 3) custom privatized variables with unknown p. We showed that a naïve debiased estimator in the first scenario results in poor MSE compared to the minimax lower bound. In contrast, we show that by using customized privacy mechanisms, we can achieve the lower bound and obtain minimax optimal inference. We also presented a Bayesian methodology and its sampling algorithm as an alternative to the frequentist methodologies. We emphasize that despite the simplicity of the Laplace and randomized response mechanisms we employ, our customized estimators attain the minimax lower bound, thereby ensuring optimality across any privacy mechanisms. Thus, the mechanism choice is of lesser concern. Additionally, our analyses can readily be extended to other mechanisms that add independent noise with a zero mean and known variance. Our Bayesian algorithm works effectively across a broad spectrum of privacy mechanisms if the privacy mechanism has a known likelihood. Finally, we validated the performance of our estimators via simulation studies and empirical analyses using real-world data.

A direction for future research is to develop an analytical framework for unbounded variables. Our framework is restricted to bounded variables due to considerations of the sensitivity of DP mechanisms.

Furthermore, the finite-sample performance of our estimators may be improved by more carefully designing the noise adding mechanisms; one may investigate using truncated-uniform-Laplace (Awan and Slavković, 2018), K-norm mechanisms (Hardt and Talwar, 2010; Awan and Slavković, 2021), or the minimax optimal noise mechanism for multivariate mean estimation (Duchi et al., 2018).

Finally, another direction of future work would be to develop methodologies for the PATE estimation in observational studies.

References

Agarwal, A. and R. Singh (2021). Causal inference with corrupted data: Measurement error, missing values, discretization, and differential privacy. *arXiv preprint arXiv:2107.02780*.

Apple, D. (2017). Learning with privacy at scale. *Apple Machine Learning Journal 1*(8).

Awan, J. and A. Slavković (2018). Differentially private uniformly most powerful tests for binomial data. *Advances in Neural Information Processing Systems 31*.

Awan, J. and A. Slavković (2021). Structure and sensitivity in differential privacy: Comparing k-norm mechanisms. *Journal of the American Statistical Association 116*(534), 935–954.

Barrera-Osorio, F., M. Bertrand, L. L. Linden, and F. Perez-Calle (2011, April). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in colombia. *American Economic Journal: Applied Economics 3*(2), 167–195.

D'Orazio, V., J. Honaker, and G. King (2015, 01). Differential privacy for social science inference. *SSRN Electronic Journal*.

Duchi, J. C., M. I. Jordan, and M. J. Wainwright (2018). Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association 113*(521), 182–201.

Dwork, C., F. McSherry, K. Nissim, and A. Smith (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer.

Erlingsson, Ú., V. Pihur, and A. Korolova (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pp. 1054–1067.

Evans, G. and G. King (2022, 2021). Statistically valid inferences from differentially private data releases, with application to the Facebook urls dataset. *Political Analysis*, 1–21.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics 2*(4), 615 – 629.

Ferrando, C., S. Wang, and D. Sheldon (2022). Parametric bootstrap for differentially private confidence intervals. In G. Camps-Valls, F. J. R. Ruiz, and I. Valera (Eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, Volume 151 of *Proceedings of Machine Learning Research*, pp. 1598–1618. PMLR.

Ferrari, S. and F. Cribari-Neto (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics 31*(7), 799–815.

Forastiere, L., F. Mealli, and T. J. VanderWeele (2016). Identification and estimation of causal mechanisms in clustered encouragement designs: Disentangling bed nets using Bayesian principal stratification. *Journal of the American Statistical Association 111*, 510–525.

Freedman, D. A. (2008). On regression adjustments in experiments with several treatments. *The Annals of Applied Statistics 2*(1), 176 – 196.

Hardt, M. and K. Talwar (2010). On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pp. 705–714.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Ishwaran, H. and L. F. James (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association 96*(453), 161–173.

Ishwaran, H. and M. Zarepour (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika 87*(2), 371–390.

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

Komarova, T. and D. Nekipelov (2020). Identification and formal privacy guarantees. *arXiv preprint arXiv:2006.14732*.

Kusner, M. J., Y. Sun, K. Sridharan, and K. Q. Weinberger (2016). Private causal inference. *International Conference on Artificial Intelligence and Statistics 51*, 1308–1317.

Lee, S. K., L. Gresele, M. Park, and K. Muandet (2019). Privacy-preserving causal inference via inverse probability weighting. *arXiv preprint arXiv:1905.12592*.

Lehmann, E. L. and G. Casella (1998). *Theory of Point Estimation* (Second ed.). New York, NY, USA: Springer-Verlag.

Lei, J., A.-S. Charest, A. Slavkovic, A. Smith, and S. Fienberg (2017, 10). Differentially Private Model Selection with Penalized and Constrained Likelihood. *Journal of the Royal Statistical Society Series A: Statistics in Society 181*(3), 609–633.

Li, F., P. Ding, and F. Mealli (2023). Bayesian causal inference: A critical review. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 381*(2247), 20220153.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics 7*(1), 295 – 318.

Maceachern, S. (1999, 01). Dependent nonparametric processes. *Proceedings of the Section on Bayesian Statistical Science, American Statistical Association*, 50–55.

Niu, F., H. Nori, B. Quistorff, R. Caruana, D. Ngwe, and A. Kannan (2022). Differentially private estimation of heterogeneous causal effects. In *First Conference on Causal Learning and Reasoning*.

Ohnishi, Y. and A. Sabbaghi (2022a). A Bayesian analysis of two-stage randomized experiments in the presence of interference, treatment nonadherence, and missing outcomes. *Bayesian Analysis*, 1 – 30.

Ohnishi, Y. and A. Sabbaghi (2022b). Degree of interference: A general framework for causal inference under interference. *arXiv preprint arXiv:2210.17516*.

Rubin, D. B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics 6*(1), 34 – 58.

Schein, A., Z. S. Wu, A. Schofield, M. Zhou, and H. Wallach (2019, 09–15 Jun). Locally private Bayesian inference for count models. In K. Chaudhuri and R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning*, Volume 97 of *Proceedings of Machine Learning Research*, pp. 5638–5648. PMLR.

Schwartz, S. L., F. Li, and F. Mealli (2011, 12). A Bayesian semiparametric approach to intermediate variables in causal inference. *Journal of the American Statistical Association 106*, 1331–1344.

Stephens, D. A., W. S. Nobre, E. E. M. Moodie, and A. M. Schmidt (2023). Causal Inference Under Mis-Specification: Adjustment Based on the Propensity Score (with Discussion). *Bayesian Analysis 18*(2), 639 – 694.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*(398), 528–540.

Zigler, C. M. (2016, March). The central role of Bayes' theorem for joint estimation of causal effects and propensity scores. *The American Statistician 70*(1), 47–54.

A Details of Theorems and Proofs in Section 3

A.1 Conditional independence of $\{Y_i(0), Y_i(1)\}$ and $\tilde{W}_i$ given W_i

We first state a subtle yet important lemma that we will use to prove subsequent theorems.

Lemma 3. *The potential outcomes are conditionally independent of the privatized treatment assignments given the actual treatment assignment:*

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp \tilde{W}_i \mid W_i.$$

This result holds because the DP mechanism flips the given treatment independently. This result is subtle, but important because it plays a crucial role in proving the upcoming theorems.

A.2 Proof of Lemma 1

Proof. We first acknowledge that

$$\sup_{\substack{P_0=\delta(0),\\ P_1\in\mathcal{P}_1,\\ p=1}} \mathbb{E}[(\hat{\tau}-\tau)^2] = \sup_{P_1\in\mathcal{P}_1} \mathbb{E}[(\hat{\tau}-\mu_1)^2], \tag{6}$$

where $\delta(0)$ denotes a point mass at 0 and $\mu_1 = \mathbb{E}[Y_i(1)]$. Equation (6) is equivalent to the one-dimensional mean estimation problem in Duchi et al. (2018, Corollary 1). Therefore, by Duchi et al. (2018), there exists some constant c_l such that

$$c_l \min(1, (N\epsilon^2)^{-1}) \leq \sup_{P_1\in\mathcal{P}_1} \mathbb{E}[(\hat{\tau}-\mu_1)^2],$$

Finally, we note that

$$\inf_{M_\epsilon\in\mathcal{M}_\epsilon} \inf_{\hat{\tau}} \sup_{\substack{P_0=\delta(0),\\ P_1\in\mathcal{P}_1,\\ p=1}} \mathbb{E}[(\hat{\tau}-\tau)^2] \leq \inf_{M_\epsilon\in\mathcal{M}_\epsilon} \inf_{\hat{\tau}} \sup_{\substack{P_0\in\mathcal{P}_0,\\ P_1\in\mathcal{P}_1,\\ p\in[0,1]}} \mathbb{E}[(\hat{\tau}-\tau)^2],$$

where the inequality holds as the right side is taking supremum over a larger set. Putting everything together, we prove our claim. □

A.3 Proof of Lemma 2

Proof. Let $\bar{p} = 1 - p$ and $\bar{q}_{\epsilon_w} = 1 - q_{\epsilon_w}$. The weak law of large numbers implies

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\tilde{W}_i\tilde{Y}_i \xrightarrow{p} \mathbb{E}[\tilde{W}_i\tilde{Y}_i]\\ &= \mathbb{E}[\mathbb{E}[\tilde{W}_i\tilde{Y}_i \mid W_i]]\\ &= \mathbb{E}[\mathbb{E}[\tilde{W}_i \mid W_i]\mathbb{E}[Y_i \mid W_i]]\\ &= \mathbb{E}[P(\tilde{W}_i = 1 \mid W_i)\mathbb{E}[\tilde{Y}_i \mid W_i]]\\ &= \mathbb{E}[P(\tilde{W}_i = 1 \mid W_i)\mathbb{E}[Y_i \mid W_i]]\\ &= p\big(P(\tilde{W}_i = 1 \mid W_i = 1)\mathbb{E}[Y_i(1)]\big) + \bar{p}\big(P(\tilde{W}_i = 1 \mid W_i = 0)\mathbb{E}[Y_i(0)]\big)\\ &= pq_{\epsilon_w}\mu_1 + \bar{p}\bar{q}_{\epsilon_w}\mu_0,\end{aligned}$$

where the second line follows from the law of total expectation and the third line follows from Lemma 3. Similarly, we have

$$\frac{1}{N}\sum_{i=1}^{N}(1 - \tilde{W}_i)\tilde{Y}_i \xrightarrow{p} p\bar{q}_{\epsilon_w}\mu_1 + \bar{p}q_{\epsilon_w}\mu_0.$$

Therefore, we see that

$$\tilde{\tau}_{naive} \xrightarrow{p} \frac{1}{C_{p,\epsilon_w}}\tau,$$

and, since C_{p,ϵ_w} is a constant, we have

$$C_{p,\epsilon_w}\tilde{\tau}_{naive} \xrightarrow{p} \tau.$$

□

A.4 Details of Theorem 3.1

We provide the following central limit theorem.

Theorem A.1. *The estimator $C_{p,\epsilon_w}\tilde{\tau}_{naive}$ is unbiased and consistent for τ. Furthermore, $\sqrt{N}(C_{p,\epsilon_w}\tilde{\tau}_{naive} - \tau)$ converges in distribution to*

$$\mathrm{N}\Bigg(0, C_{p,\epsilon_w}^2\bigg(\frac{1}{\rho_1}V_1 + \frac{1}{\rho_0}V_0 + \frac{\rho_0}{\rho_1}E_1^2 + \frac{\rho_1}{\rho_0}E_0^2 + 2E_0E_1\bigg)\Bigg), \tag{7}$$

where, for $w = 0, 1$,

$$\begin{aligned}V_w = \mathbb{V}\mathrm{ar}(\tilde{Y}_i|\tilde{W}_i = w) =& P(W_i = 0|\tilde{W}_i = w)\mathbb{V}\mathrm{ar}[Y_i(0)] + P(W_i = 1|\tilde{W}_i = w)\mathbb{V}\mathrm{ar}[Y_i(1)]\\ &+ P(W_i = 0|\tilde{W}_i = w)P(W_i = 1|\tilde{W}_i = w)\tau^2 + \frac{2}{\epsilon_y^2},\end{aligned}$$

and $E_w = \mathbb{E}(\tilde{Y}_i | \tilde{W}_i = w) = P(W_i = 0 | \tilde{W}_i = w)\mathbb{E}[Y_i(0)] + P(W_i = 1 | \tilde{W}_i = w)\mathbb{E}[Y_i(1)]$.

Proof. Consistency is proven in Section A.3.

$$C_{p,\epsilon_w}\tilde{\tau}_{naive} = \frac{C_{p,\epsilon_w}}{N}\sum_{i=1}^{N}\left\{\frac{\tilde{W}_i\tilde{Y}_i}{\rho_1} - \frac{(1-\tilde{W}_i)\tilde{Y}_i}{\rho_0}\right\} = \frac{C_{p,\epsilon_w}}{N}\sum_{i=1}^{N}\tilde{\tau}_i.$$

Note that $\tilde{\tau}_i$ is i.i.d. for $i = 1, \ldots, N$, $\mathbb{E}[C_{p,\epsilon_w}\tilde{\tau}_i] = \tau$, and the second moment is bounded due to the sensitivity of Y. Thus, it is sufficient to derive the variance of $\tilde{\tau}_i$ as $\mathbb{V}\text{ar}[C_{p,\epsilon_w}\tilde{\tau}_i] = C^2_{p,\epsilon_w}\mathbb{V}\text{ar}[\tilde{\tau}_i]$.

$$\mathbb{V}\text{ar}[\tilde{\tau}_i] = \frac{1}{\rho_1^2}\mathbb{V}\text{ar}[\tilde{W}_i\tilde{Y}_i] + \frac{1}{\rho_0^2}\mathbb{V}\text{ar}[(1-\tilde{W}_i)\tilde{Y}_i] - \frac{2}{\rho_0\rho_1}\mathbb{C}\text{ov}[\tilde{W}_i\tilde{Y}_i, (1-\tilde{W}_i)\tilde{Y}_i].$$

Then,

$$\begin{aligned}
\mathbb{V}\text{ar}[\tilde{W}_i\tilde{Y}_i] &= \mathbb{E}[\mathbb{V}\text{ar}[\tilde{W}_i\tilde{Y}_i \mid \tilde{W}_i]] + \mathbb{V}\text{ar}[\mathbb{E}[\tilde{W}_i\tilde{Y}_i \mid \tilde{W}_i]] \\
&= \mathbb{E}[\tilde{W}_i^2\mathbb{V}\text{ar}[\tilde{Y}_i \mid \tilde{W}_i]] + \mathbb{V}\text{ar}[\tilde{W}_i\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i]] \\
&= p(\tilde{W}_i = 1)\mathbb{V}\text{ar}[\tilde{Y}_i \mid \tilde{W}_i] + p(\tilde{W}_i = 1)p(\tilde{W}_i = 0)\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i]^2 \\
&= \rho_1\mathbb{V}\text{ar}[\tilde{Y}_i \mid \tilde{W}_i] + \rho_0\rho_1\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i]^2 \\
&= \rho_1 V_1 + \rho_0\rho_1 E_1^2.
\end{aligned}$$

Similarly, we have $\mathbb{V}\text{ar}[(1-\tilde{W}_i)\tilde{Y}_i] = \rho_0 V_0 + \rho_0\rho_1 E_0^2$. The covariance is given by

$$\begin{aligned}
\mathbb{C}\text{ov}[\tilde{W}_i\tilde{Y}_i, (1-\tilde{W}_i)\tilde{Y}_i] &= -\mathbb{E}[\tilde{W}_i\tilde{Y}_i]\mathbb{E}[(1-\tilde{W}_i)\tilde{Y}_i] \\
&= -\mathbb{E}[\tilde{W}_i\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i]]\mathbb{E}[(1-\tilde{W}_i)\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i]] \\
&= -p(\tilde{W}_i = 1)\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1]p(\tilde{W}_i = 0)\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 0] \\
&= -\rho_0\rho_1 E_0 E_1.
\end{aligned}$$

Putting all together, we prove the central limit theorem in Theorem A.1 and hence Theorem 3.1.

Next, we consider the decompositions of E_w and V_w. We have

$$E_w = \mathbb{E}[\hat{Y}_i \mid \hat{W}_i = w] = \mathbb{E}[Y_i \mid \tilde{W}_i = w] = P(W_i = 0 | \tilde{W}_i = w)\mathbb{E}[Y_i(0)] + P(W_i = 1 | \tilde{W}_i = w)\mathbb{E}[Y_i(1)],$$

which follows from Lemma 3. By the law of total variance and SUTVA,

$$\mathbb{V}\text{ar}[Y_i \mid \tilde{W}_i = 1] = \sum_{w=0}^{1}\mathbb{V}\text{ar}[Y_i \mid \tilde{W}_i = 1, W_i = w]P(W_i = w \mid \tilde{W}_i = 1) + \mathbb{V}\text{ar}[\mathbb{E}[Y_i \mid \tilde{W}_i = 1, W_i = w]].$$

The first term simplifies to

$$\sum_{w=0}^{1}\mathbb{V}\text{ar}[Y_i \mid \tilde{W}_i = 1, W_i = w]P(W_i = w \mid \tilde{W}_i = 1) = \frac{\bar{p}\bar{q}_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}}\mathbb{V}\text{ar}[Y_i(0)] + \frac{pq_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}}\mathbb{V}\text{ar}[Y_i(1)].$$

The second term simplifies to

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}[\mathbb{E}[Y_i \mid \tilde{W}_i = 1, W_i = w]] \\
&= \mathbb{E}\big[(\mathbb{E}[Y_i \mid \tilde{W}_i = 1, W_i = w] - \mathbb{E}[(\mathbb{E}[Y_i \mid \tilde{W}_i = 1, W_i = w]])^2 \mid \tilde{W}_i = 1\big] \\
&= \sum_{w=0}^{1} \left(\mathbb{E}[Y_i(w)] - \frac{\bar{p}\bar{q}_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}} \mathbb{E}[Y_i(0)] - \frac{pq_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}} \mathbb{E}[Y_i(1)] \right)^2 P(W_i = w \mid \tilde{W}_i = 1) \\
&= \frac{pq_{\epsilon_w}\bar{p}\bar{q}_{\epsilon_w}}{(pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w})^2} \tau^2.
\end{aligned}$$

Therefore, we have

$$V_1 := \mathbb{V}\mathrm{ar}[Y_i \mid \tilde{W}_i = 1] = \frac{\bar{p}\bar{q}_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}} \mathbb{V}\mathrm{ar}[Y_i(0)] + \frac{pq_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}} \mathbb{V}\mathrm{ar}[Y_i(1)] + \frac{pq_{\epsilon_w}\bar{p}\bar{q}_{\epsilon_w}}{(pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w})^2} \tau^2.$$

Similarly, we have

$$V_0 := \mathbb{V}\mathrm{ar}[Y_i \mid \tilde{W}_i = 0] = \frac{\bar{p}q_{\epsilon_w}}{\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w}} \mathbb{V}\mathrm{ar}[Y_i(0)] + \frac{p\bar{q}_{\epsilon_w}}{\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w}} \mathbb{V}\mathrm{ar}[Y_i(1)] + \frac{pq_{\epsilon_w}\bar{p}\bar{q}_{\epsilon_w}}{(\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w})^2} \tau^2.$$

Finally, the order of the asymptotic variance is immediate from the fact that $C^2_{p,\epsilon_w} = O((\epsilon_w^2)^{-1})$, which proves Theorem 3.1 and Corollary 1

□

We now turn to estimating the asymptotic variance of $C_{p,\epsilon_w}\tilde{\tau}_{naive}$ in (7). We consider the following estimators for E_w and V_w: $\hat{E}_w = \frac{1}{\tilde{N}_w} \sum_{i:\tilde{W}_i = w} \tilde{Y}_i$ and $\hat{V}_w = \frac{1}{\tilde{N}_w - 1} \sum_{i:\tilde{W}_i = w} (\tilde{Y}_i - \hat{E}_w)^2$, where $\tilde{N}_w = \sum_{i=1}^{N} \mathbb{1}(\tilde{W}_i = w)$ for $w = 0, 1$.

Lemma 4. *$\hat{V}_w$ and $\hat{E}_w$ are consistent for V_w and E_w respectively. Also, we have*

$$\mathbb{E}[\hat{E}_w \mid \tilde{W}_i = w] = E_w \text{ and } \mathbb{E}[\hat{V}_w \mid \tilde{W}_i = w] = V_w$$

Proof.

$$\begin{aligned}
\hat{V}_1 &= \frac{1}{\tilde{N}_1 - 1} \sum_{i:\tilde{W}_i = 1} (\tilde{Y}_i - \hat{E}_1)^2 \\
&= \frac{1}{\tilde{N}_1 - 1} \sum_{i:\tilde{W}_i = 1} (\tilde{Y}_i - \mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1] + \mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1] - \hat{E}_1)^2 \\
&= \frac{1}{\tilde{N}_1 - 1} \sum_{i:\tilde{W}_i = 1} \Big\{ (\tilde{Y}_i - \mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1])^2 + (\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1] - \hat{E}_1)^2 \\
&\quad - 2(\tilde{Y}_i - \mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1])(\mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1] - \hat{E}_1) \Big\} \\
&= \frac{\tilde{N}_1}{\tilde{N}_1 - 1} \frac{1}{\tilde{N}_1} \sum_{i:\tilde{W}_i = 1} (\tilde{Y}_i - \mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1])^2 - \frac{\tilde{N}_1}{\tilde{N}_1 - 1} (\hat{E}_1 - \mathbb{E}[\tilde{Y}_i \mid \tilde{W}_i = 1])^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}[\hat{V}_1 \mid \tilde{W}_i = 1] &= \frac{\tilde{N}_1}{\tilde{N}_1 - 1}\mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] - \frac{\tilde{N}_1}{\tilde{N}_1 - 1}\mathbb{V}\mathrm{ar}[\hat{E}_1 \mid \tilde{W}_i = 1] \\
&= \frac{\tilde{N}_1}{\tilde{N}_1 - 1}\mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] - \frac{1}{\tilde{N}_1 - 1}\mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] \\
&= \mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] \\
&= V_1.
\end{aligned}$$

We can follow the same procedure for $\mathbb{E}[\hat{V}_0 \mid \tilde{W}_i = 0] = V_0$. □

Using $\hat{E}_w$ and $\hat{V}_w$, we can construct the plug-in estimator for the asymptotic variance and the nominal central confidence interval at the significance level α as:

$$\left(C_{p,\epsilon_w}\tilde{\tau}_{naive} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{naive}}{N}}, C_{p,\epsilon_w}\tilde{\tau}_{naive} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{naive}}{N}}\right).$$

where $\hat{\Sigma}_{naive} = C_{p,\epsilon_w}^2(\frac{1}{\rho_1}\hat{V}_1 + \frac{1}{\rho_0}\hat{V}_0 + \frac{\rho_0}{\rho_1}\hat{E}_1^2 + \frac{\rho_1}{\rho_0}\hat{E}_0^2 + 2\hat{E}_0\hat{E}_1)$, which is a consistent estimator for the asymptotic variance in (7).

Finally, we discuss the optimality of the naïve estimator.

Corollary 1 (Convergence rate)**.** *The naïve estimator under the joint scenario has the MSE $O((N\epsilon_y^2\epsilon_w^2)^{-1})$.*

Setting $\epsilon_y = \epsilon_2 = \epsilon/2$ gives $O((N\epsilon^4)^{-1})$. While we do not match the minimax lower bound of mean estimation in terms of ϵ when both W and Y are privatized, it should be emphasized that the estimation of PATE is significantly harder than the usual mean estimation when we do not know who belongs to which treatment group, especially using a non-interactive LDP mechanism as in the joint scenario.

A.5 Details of Theorem 3.2

By the standard central limit theorem, we have

$$\sqrt{N}(\tilde{\tau}_{IPW} - \tau) \xrightarrow{D} \mathrm{N}\left(0, \frac{\mu_1^2 + \sigma_1^2}{p} + \frac{\mu_0^2 + \sigma_0^2}{1-p} - \tau^2 - \mu_0\mu_1 + \frac{2\Delta_A}{\epsilon_a^2}\right), \quad (8)$$

where $\mu_w = \mathbb{E}[Y_i(w)]$ and $\sigma_w^2 = \mathbb{V}\mathrm{ar}[Y_i(w)]$ for $w = 0, 1$. We can then construct the plug-in estimator for the asymptotic variance and the nominal central confidence interval at the significance level α as:

$$\left(\tilde{\tau}_{IPW} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{IPW}}{N}}, \tilde{\tau}_{IPW} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{IPW}}{N}}\right).$$

where $\hat{\Sigma}_{IPW} = \frac{1}{N-1}\sum_{i=1}^N(\tilde{A}_i - \hat{E}_A)^2$ with $\hat{E}_A = \frac{1}{N}\sum_{i=1}^N \tilde{A}_i$, which is an unbiased estimator for the asymptotic variance in (8).

A.6 Details of Theorem 3.3

First, we provide the following asymptotic results regarding this estimator.

Theorem A.2. *$\tilde{\tau}_{DM}$ is consistent for τ and $\sqrt{N}(\tilde{\tau}_{DM} - \tau)$ converges in distribution to*

$$\mathrm{N}\left(0, 4\mu_0\mu_1 + \frac{\sigma_0^2}{1-p} + \frac{\sigma_1^2}{p} + \frac{2}{\epsilon_{b_1}^2}\left(\frac{\mu_0}{1-p} + \frac{\mu_1}{p}\right)^2 + \frac{2}{p^2\epsilon_{b_2}^2} + \frac{2}{(1-p)^2\epsilon_{b_3}^2}\right). \tag{9}$$

Proof. First, we have

$$\mathbb{E}[\tilde{B}_{i,1}] = p\mu_1, \mathbb{E}[\tilde{B}_{i,2}] = (1-p)\mu_0, \mathbb{E}[\tilde{B}_{i,3}] = p, \mathbb{E}[\tilde{B}_{i,4}] = 1-p, \mathbb{V}\mathrm{ar}[\tilde{B}_{i,1}] = p\sigma_1^2 + p(1-p)\mu_1^2 + \frac{2}{\epsilon_{b1}^2},$$
$$\mathbb{V}\mathrm{ar}[\tilde{B}_{i,2}] = (1-p)\sigma_0^2 + p(1-p)\mu_0^2 + \frac{2}{\epsilon_{b2}^2}, \mathbb{V}\mathrm{ar}[\tilde{B}_{i,3}] = p(1-p) + \frac{2}{\epsilon_{b3}^2}, \mathbb{V}\mathrm{ar}[\tilde{B}_{i,4}] = p(1-p) + \frac{2}{\epsilon_{b3}^2},$$
$$\mathbb{C}\mathrm{ov}[\tilde{B}_{i,1}, \tilde{B}_{i,2}] = -p(1-p)\mu_0\mu_1, \mathbb{C}\mathrm{ov}[\tilde{B}_{i,1}, \tilde{B}_{i,3}] = p(1-p)\mu_1, \mathbb{C}\mathrm{ov}[\tilde{B}_{i,1}, \tilde{B}_{i,4}] = 0,$$
$$\mathbb{C}\mathrm{ov}[\tilde{B}_{i,2}, \tilde{B}_{i,3}] = 0, \mathbb{C}\mathrm{ov}[\tilde{B}_{i,2}, \tilde{B}_{i,4}] = p(1-p)\mu_0, \mathbb{C}\mathrm{ov}[\tilde{B}_{i,3}, \tilde{B}_{i,4}] = -p(1-p)\mu_0\mu_1.$$

By the central limit theorem, we have

$$\sqrt{N}\begin{pmatrix} \frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,1} - \mathbb{E}[\tilde{B}_{i,1}] \\ \frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,2} - \mathbb{E}[\tilde{B}_{i,2}] \\ \frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,3} - \mathbb{E}[\tilde{B}_{i,3}] \\ \frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,4} - \mathbb{E}[\tilde{B}_{i,4}] \end{pmatrix} \xrightarrow{D} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, S^*\right),$$

where

$$S^* = \begin{pmatrix} \mathbb{V}\mathrm{ar}[\tilde{B}_{i,1}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,1}, \tilde{B}_{i,2}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,1}, \tilde{B}_{i,3}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,1}, \tilde{B}_{i,4}] \\ \mathbb{C}\mathrm{ov}[\tilde{B}_{i,2}, \tilde{B}_{i,1}] & \mathbb{V}\mathrm{ar}[\tilde{B}_{i,2}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,2}, \tilde{B}_{i,3}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,2}, \tilde{B}_{i,4}] \\ \mathbb{C}\mathrm{ov}[\tilde{B}_{i,3}, \tilde{B}_{i,1}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,3}, \tilde{B}_{i,2}] & \mathbb{V}\mathrm{ar}[\tilde{B}_{i,3}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,3}, \tilde{B}_{i,4}] \\ \mathbb{C}\mathrm{ov}[\tilde{B}_{i,4}, \tilde{B}_{i,1}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,4}, \tilde{B}_{i,2}] & \mathbb{C}\mathrm{ov}[\tilde{B}_{i,4}, \tilde{B}_{i,3}] & \mathbb{V}\mathrm{ar}[\tilde{B}_{i,4}] \end{pmatrix}.$$

Define a function $h(a,b,c,d) = \frac{a}{c} - \frac{b}{d}$ and $\nabla h = (\frac{\partial h}{\partial a}, \frac{\partial h}{\partial b}, \frac{\partial h}{\partial c}, \frac{\partial h}{\partial d})$, where

$$\frac{\partial h}{\partial a} = \frac{1}{c}, \frac{\partial h}{\partial b} = -\frac{1}{d}, \frac{\partial h}{\partial c} = -\frac{a}{c^2}, \frac{\partial h}{\partial d} = \frac{b}{d^2}.$$

Note that

$$\tau = \mu_1 - \mu_0 = \frac{\mathbb{E}[\tilde{B}_{i,1}]}{\mathbb{E}[\tilde{B}_{i,3}]} - \frac{\mathbb{E}[\tilde{B}_{i,2}]}{\mathbb{E}[\tilde{B}_{i,4}]} = h\left(\mathbb{E}[\tilde{B}_{i,1}], \mathbb{E}[\tilde{B}_{i,2}], \mathbb{E}[\tilde{B}_{i,3}], \mathbb{E}[\tilde{B}_{i,4}]\right),$$

and

$$\tilde{\tau}_{DM} = \frac{\sum_{i=1}^N \tilde{B}_{i,1}}{\sum_{i=1}^N \tilde{B}_{i,3}} - \frac{\sum_{i=1}^N \tilde{B}_{i,2}}{\sum_{i=1}^N \tilde{B}_{i,4}} = h\left(\frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,1}, \frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,2}, \frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,3}, \frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,4}\right).$$

By applying the delta method, we have

$$\sqrt{N}(\tilde{\tau}_{DM} - \tau) \xrightarrow{D} \mathrm{N}(0, \Sigma^*),$$

where $\Sigma^* = \nabla h(\mathbf{E})' S^* \nabla h(\mathbf{E})$. $\nabla h(\mathbf{E})$ denotes ∇h evaluated at $\mathbf{E} = (\mathbb{E}[\tilde{B}_{i,1}], \mathbb{E}[\tilde{B}_{i,2}, \mathbb{E}[\tilde{B}_{i,3}], \mathbb{E}[\tilde{B}_{i,4}])$. Calculating Σ^* proves our claim in Thereom A.2. The estimator of Σ^* that we adopt in Section 3.4 are a plug-in estimator with consistent estimators of $\nabla h(\mathbf{E})$ and S^*.

□

By Theorem A.2, the asymptotic variance of $\tilde{\tau}_{DM}$ has the convergence rate $O((N(\epsilon_{b_1}^2 + \epsilon_{b_2}^2 + \epsilon_{b_3}^2))^{-1})$. Setting $\epsilon_{b_1} = \epsilon_{b_2} = \epsilon_{b_3} = \epsilon/3$ gives $O((N\epsilon^2)^{-1})$, which also matches the minimax lower bound for the locally private mean estimation, indicating the optimality of the estimator.

Let $\hat{E}_{B_j} = \frac{1}{N}\sum_{i=1}^N \tilde{B}_{i,j}$, $\hat{V}_{B_j} = \frac{1}{N-1}\sum_{i=1}^N (\tilde{B}_{i,j} - \hat{E}_{B_j})^2$ for $j = 1, 2, 3, 4$ and $\widehat{\mathrm{Cov}_{j,k}} = \frac{1}{N-1}\sum_{i=1}^N (\tilde{B}_{i,j} - \hat{E}_{B_j})(\tilde{B}_{i,k} - \hat{E}_{B_k})$ for $j \neq k$. Then, we construct the plug-in estimator for the asymptotic variance and the nominal central confidence interval at the significance level α as:

$$\left(\tilde{\tau}_{DM} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{DM}}{N}}, \tilde{\tau}_{DM} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{DM}}{N}}\right).$$

where $\hat{\Sigma}_{DM} = \hat{\mathbf{e}}'\hat{\mathbf{S}}\hat{\mathbf{e}}$, with $\hat{\mathbf{e}} = (1/\hat{E}_{B_3}, -1/(1-\hat{E}_{B_3}), -\hat{E}_{B_1}/\hat{E}_{B_3}^2, \hat{E}_{B_2}/(1-\hat{E}_{B_3})^2)'$ and

$$\hat{\mathbf{S}} = \begin{pmatrix} \hat{V}_{B_1} & \widehat{\mathrm{Cov}_{1,2}} & \widehat{\mathrm{Cov}_{1,3}} & \widehat{\mathrm{Cov}_{1,4}} \\ \widehat{\mathrm{Cov}_{2,1}} & \hat{V}_{B_2} & \widehat{\mathrm{Cov}_{2,3}} & \widehat{\mathrm{Cov}_{2,4}} \\ \widehat{\mathrm{Cov}_{3,1}} & \widehat{\mathrm{Cov}_{3,2}} & \hat{V}_{B_3} & \widehat{\mathrm{Cov}_{3,4}} \\ \widehat{\mathrm{Cov}_{4,1}} & \widehat{\mathrm{Cov}_{4,2}} & \widehat{\mathrm{Cov}_{4,3}} & \hat{V}_{B_4} \end{pmatrix}.$$

This is a consistent estimator for the asymptotic variance in (9).

B Bayesian Methodology

B.1 Details of the DPM

We say the probability measure H is generated from a Dirichlet Process, $\mathrm{DP}(\alpha, H_0)$, with a concentration parameter $\alpha > 0$ and a base probability measure H_0 over a measurable space $(\Theta, \mathcal{S})$ (Ferguson, 1974) if, for any finite partition $(S_1, ..., S_k)$ of $\mathcal{S}$, we have

$$\big(H(S_1), ..., H(S_k)\big) \sim \mathrm{Dir}\big(\alpha H_0(S_1), ..., \alpha H_0(S_k)\big),$$

where $\mathrm{Dir}(\alpha_1, ..., \alpha_k)$ denotes the Dirichlet distribution with positive parameters $\alpha_1, ..., \alpha_k$. The DPM is specified as

$$\begin{aligned}\{Y_1(0), Y_1(1)\}, ..., \{Y_N(0), Y_N(1)\} \mid \Phi_1, ..., \Phi_N &\overset{ind}{\sim} p(Y_i(0), Y_i(1)|\Phi_i),\\ \Phi_1, ..., \Phi_N|H &\overset{ind}{\sim} H,\\ H &\overset{ind}{\sim} DP(\alpha, H_0).\end{aligned}$$

We write $\overset{ind}{\sim}$ to say *independently distributed*. This model has unit-level parameters Φ_i for $i = 1, ..., N$, but the discreteness of the Dirichlet process (DP) distributed prior implies that the vector $\boldsymbol{\Phi} = (\Phi_1, ..., \Phi_N)$ can be rewritten in terms of its unique values $\boldsymbol{\Phi}^* = (\Phi_1^*, ..., \Phi_K^*)$. In particular, this can be represented in the following stick-breaking process.

$$H = \sum_{k=1}^{\infty} u_k \delta_{\Phi_k}, \;\; u_k = v_k \prod_{l<k} [1 - v_l], \;\; v_l \overset{ind}{\sim} \mathrm{Beta}(1, \alpha).$$

More specifically, the outcome model is specified by the following model.

$$P(Y_i(w)|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \sum_{k=1}^{\infty} u_k \mathrm{TN}(\mu_w^k, \Sigma_w^k, 0, 1), \tag{10}$$

where $\mathrm{TN}(\mu, \sigma^2, u, l)$ denotes the truncated normal distribution with the mean, variance, upper bound and lower bound parameters. The atoms $\Phi_k = (\mu_0^k, \mu_1^k, \Sigma_0^k, \Sigma_1^k)$ and the weight parameters u_k are nonparametrically specified via $\mathrm{DP}(\alpha, H_0)$. This can be regarded as the infinite mixture of normal distributions, where μ_w^k and Σ_w^k is the location parameter and variance parameter of each component respectively.

For inference, we adopt an approximated blocked Gibbs sampler based on a truncation of the stick-breaking representation of the DP proposed by Ishwaran and Zarepour (2000), due to its simplicity. In this algorithm, we first set a conservatively large upper bound, $K \leq \infty$, on the number of components that units potentially belong to. Let $C_i \in \{1, ..., K\}$ denote the latent class indicators with a multinomial distribution, $C_i \sim MN(\mathbf{w})$ where $\mathbf{u} = (u_1, ..., u_K)$ denote the weights of all components of the DPM. Conditional on $C_i = k$, (10) is greatly simplified to

$$P(Y_i(w)|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \mathrm{TN}(\mu_w^k, \Sigma_w^k, 0, 1).$$

Ishwaran and James (2001) showed that an accurate approximation to the exact DP is obtained as long as K is chosen sufficiently large. The DPM provides an automatic selection mechanism for the number of active components $K^* < K$. To ensure that K is sufficiently large, we run several MCMC iterations with different values of K. If the current iteration occupies all components, then K is not large enough, so we increase K for the next iteration. We conduct this iterative process until the number of the occupied components is below K.

B.2 Detailed Steps of Gibbs Sampler

In this section we present the detailed steps of the Gibbs sampler that is described in Section 4.2. The algorithm is inspired by Schwartz et al. (2011) and Ohnishi and Sabbaghi (2022b).

1. Given $Y_i(0), Y_i(1)$, draw each W_i from

$$P(W_i = 1|-) = \frac{r_1}{r_0 + r_1},$$

where, for unit i with $\tilde{W}_i = 0$,

$$r_0 = \mathrm{Lap}(\tilde{\mathrm{Y}}_\mathrm{i} \mid \mathrm{Y_i}(0), 1/\epsilon_\mathrm{y}) q_{\epsilon_w}(1-p) \text{ and } r_1 = \mathrm{Lap}(\tilde{\mathrm{Y}}_\mathrm{i} \mid \mathrm{Y_i}(1), 1/\epsilon_\mathrm{y})(1 - q_{\epsilon_w})p,$$

and for unit i with $\tilde{W}_i = 1$,

$$r_0 = \mathrm{Lap}(\tilde{\mathrm{Y}}_\mathrm{i} \mid \mathrm{Y_i}(0), 1/\epsilon_\mathrm{y})(1 - q_{\epsilon_w})(1-p) \text{ and } r_1 = \mathrm{Lap}(\tilde{\mathrm{Y}}_\mathrm{i} \mid \mathrm{Y_i}(1), 1/\epsilon_\mathrm{y}) q_{\epsilon_w} p.$$

where $\mathrm{Lap}(\mathrm{y} \mid \mu, \sigma)$ is the pdf of the laplace distribution evaluated at y with the location parameter μ and scale parameter σ.

2. Given $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{u}$, C_i and $W_i = w$, draw $Y_i(1-w)$ according to:

$$Y_i(1-w) \sim \mathrm{TN}(\mu_{1-w}^{C_i}, \Sigma_{1-w}^{C_i}, 0, 1),$$

where $\mathrm{TN}(\mu, \sigma^2, u, l)$ denotes the truncated normal distribution with the mean, variance, upper bound and lower bound parameters.
Then, draw $Y_i(w)$ using the following Privacy-Aware Metropolis-within-Gibbs sampler Ju et al. (2022):
   (a) Draw a proposal: $y* \sim \mathrm{TN}(\mu_w^{C_i}, \Sigma_w^{C_i}, 0, 1)$.
   (b) Accept the proposal with probability $\alpha = \min\left(1, \frac{\mathrm{Lap}(\mathrm{y}* \mid \tilde{\mathrm{Y}}_\mathrm{i}, 1/\epsilon_\mathrm{y})}{\mathrm{Lap}(\mathrm{y}^{\mathrm{prev}} \mid \tilde{\mathrm{Y}}_\mathrm{i}, 1/\epsilon_\mathrm{y})}\right)$,
   where y^{prev} is the value of $Y_i(w)$ in the previous step.

3. Given $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{u}$, $Y_i(0)$ and $Y_i(1)$, draw each C_i from

$$P(C_i = k|-) \propto u_k \mathrm{TN}(Y_i(0) \mid \mu_0^k, \Sigma_0^k, 0, 1) \mathrm{TN}(Y_i(1) \mid \mu_1^k, \Sigma_1^k, 0, 1).$$

This is a multinomial distribution.

4. Let $u'_K = 1$. Given α, $\mathbf{C}$, draw u'_k for $k \in \{1, ..., K-1\}$ from

$$P(u'_k|-) \propto \mathrm{Beta}\left(1 + \sum_{i:C_i=k} 1, \alpha + \sum_{i:C_i>k} 1\right).$$

Then, update $u_k = u'_k \prod_{j<k}(1 - u'_j)$.

5. Given $\mathbf{C}$ and $\mathbf{u}'$, draw α from

$$P(\alpha|-) \propto P(\alpha) \prod_{k=1}^{K} f\left(u'_k \middle| 1 + \sum_{i:C_i=k} 1, \alpha + \sum_{i:C_i>k} 1\right),$$

where f is the pdf of u'_k, the beta distribution. The Metropolis-Hastings algorithm is used for this step with a proposal distribution $\mathrm{TN}(\alpha^{prev}, 1.0, 0, \infty)$. α^{prev} is the value of α in the previous step.

6. Given $\mathbf{Y}(0)$, $\mathbf{Y}(1)$ and $\mathbf{C}$, draw $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from
   (a) If $N_k = \sum_{i=1}^{N} \mathbb{1}(C_i = k) > 0$, draw Σ_w^k from $\mathrm{IG}(2 + 0.5N_k, 0.2^2 + 0.5s_w^k)$ where $s_w^k = \sum_{i:C_i=k}(Y_i(w) - \mu_w^k)^2$ for $w = 0, 1$. If $N_k = 0$,then draw Σ_w^k from the prior $\mathrm{IG}(2, 0.2^2)$.
   (b) If $N_k > 0$, draw μ_w^k from

$$\mathrm{TN}\left(\frac{0.5 * \Sigma_w^k + 9.0s_w}{\Sigma_w^k + 9.0N_k}, \frac{9.0\Sigma_w^k}{\Sigma_w^k + 9.0N_k}, 0, 1\right),$$

   where $s_w = \sum_{i=1}^{N} Y_i(w)$. If $N_k = 0$, draw μ_w^k from

$$\mathrm{TN}(0.5, 9.0, 0, 1).$$

   We use a common choice of the base measure H_0: the Normal-Inverse-Gamma conjugate $\mathrm{N}(\mu_0, \sigma_0^2)\mathrm{N}(\mu_0, \sigma_0^2)\mathrm{IG}(a_0, b_0)\mathrm{IG}(a_0, b_0)$. The specific values of the hyperparameters in this step are: $\mu_0 = 0.5$, $\sigma_0 = 3.0$, $a_0 = 2.0$ and $b_0 = 0.2^2$ for both $w = 0, 1$.

B.3 Modifications for Custom Scenario in Section 3.3

We need to modify Step 1 and 2 for the custom scenarios. Particularly,

1. Given $Y_i(0), Y_i(1)$, draw each W_i from

$$P(W_i = 1|-) = \frac{r_1}{r_0 + r_1},$$

   where $r_w = P(\tilde{A}_i \mid Y_i(0), Y_i(1), W_i = w)P(W_i = w)$ for $w = 0, 1$. Specifically, since $\tilde{A}_i$ is generated by privatizing either $-Y_i(0)/(1-p)$ or $Y_i(1)/p$ given the value of W_i, $P(\tilde{A}_i \mid Y_i(0), Y_i(1), W_i = w) = \mathrm{Lap}(\tilde{A}_i \mid -Y_i(0)/(1-p), \Delta_a/\epsilon_a)$ for $W_i = 0$, and $P(\tilde{A}_i \mid Y_i(0), Y_i(1), W_i = w) = \mathrm{Lap}(\tilde{A}_i \mid Y_i(1)/p, \Delta_a/\epsilon_a)$ for $W_i = 1$.
2. Given $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{u}$, C_i and W_i, draw each $Y_i(0)$ and $Y_i(1)$ according to:

$$P(Y_i(W_i)|-) \propto P(Y_i(W_i) \mid \mu_{W_i}^{C_i}, \Sigma_{W_i}^{C_i})P(\tilde{A}_i \mid Y_i(W_i))$$
$$P(Y_i(1-W_i)|-) \propto P(Y_i(1-W_i) \mid \mu_{1-W_i}^{C_i}, \Sigma_{1-W_i}^{C_i}).$$

   Specifically, $P(\tilde{A}_i \mid Y_i(W_i)) = \mathrm{Lap}(\tilde{A}_i \mid -Y_i(0)/(1-p), \Delta_a/\epsilon_a)$ for $W_i = 0$ and $P(\tilde{A}_i \mid Y_i(W_i)) = \mathrm{Lap}(\tilde{A}_i \mid Y_i(1)/p, \Delta_a/\epsilon_a)$ for $W_i = 1$. The privacy-aware Metropolis-within-Gibbs algorithm (Ju et al., 2022) is used for the draw of $Y_i(W_i)$.

B.4 Modifications for Custom Scenario in Section 3.4

Under the custom scenario in Section 3.4, we do not have access to p. Therefore, we need an additional step to infer p. Specifically, with a prior distribution $p \sim \mathrm{Beta}(1, 1)$, we add the following step.

0. Draw $p \sim \mathrm{Beta}\left(1 + \sum_{i=1}^{N} \mathbb{1}(W_i = 1), 1 + \sum_{i=1}^{N} \mathbb{1}(W_i = 0)\right)$.

Then we proceed as follows.

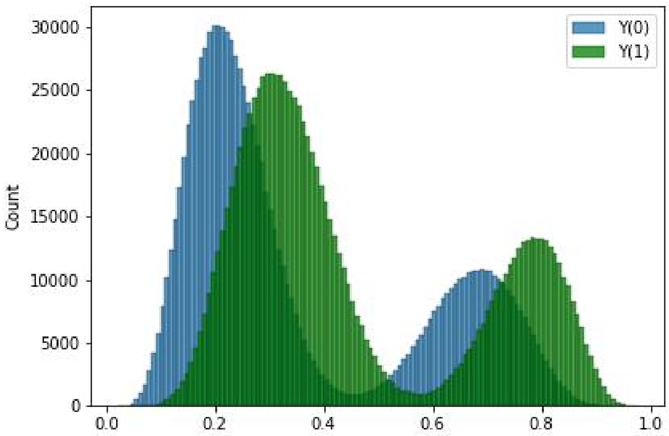

Figure 1: Distributions of $Y(0)$ and $Y(1)$ for simulation studies.

1. Given $Y_i(0), Y_i(1)$ and p, draw each W_i from

$$P(W_i = 1|-) = \frac{r_1}{r_0 + r_1},$$

where

$$r_w = p^w(1-p)^{1-w} P(\tilde{B}_{i,1} \mid Y_i(0), Y_i(1), W_i = w) P(\tilde{B}_{i,2} \mid Y_i(0), Y_i(1), W_i = w) \\ \times P(\tilde{B}_{i,3} \mid Y_i(0), Y_i(1), W_i = w)$$

for $w = 0, 1$. Specifically, considering the privatization of $\tilde{B}_{i,1}$, $\tilde{B}_{i,2}$ and $\tilde{B}_{i,3}$, we have $P(\tilde{B}_{i,1} \mid Y_i(0), Y_i(1), W_i = 0) = \text{Lap}(\tilde{B}_{i,1} \mid 0, 1/\epsilon_{b_2})$, $P(\tilde{B}_{i,2} \mid Y_i(0), Y_i(1), W_i = 0) = \text{Lap}(\tilde{B}_{i,2} \mid Y_i(0), 1/\epsilon_{b_2})$, $P(\tilde{B}_{i,3} \mid Y_i(0), Y_i(1), W_i = 0) = \text{Lap}(\tilde{B}_{i,3} \mid 0, 1/\epsilon_{b_3})$, $P(\tilde{B}_{i,1} \mid Y_i(0), Y_i(1), W_i = 1) = \text{Lap}(\tilde{B}_{i,1} \mid Y_i(1), 1/\epsilon_{b_1})$, $P(\tilde{B}_{i,2} \mid Y_i(0), Y_i(1), W_i = 1) = \text{Lap}(\tilde{B}_{i,2} \mid 0, 1/\epsilon_{b_2})$ and $P(\tilde{B}_{i,3} \mid Y_i(0), Y_i(1), W_i = 1) = \text{Lap}(\tilde{B}_{i,3} \mid 0, 1/\epsilon_{b_3})$.
2. Given $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{u}$, C_i and W_i, draw each $Y_i(0)$ and $Y_i(1)$ according to:

$$P(Y_i(W_i)|-) \propto P(Y_i(W_i) \mid \mu_{W_i}^{C_i}, \Sigma_{W_i}^{C_i}) P(\tilde{\mathbf{B}}_i \mid Y_i(W_i)) \\ P(Y_i(1-W_i)|-) \propto P(Y_i(1-W_i) \mid \mu_{1-W_i}^{C_i}, \Sigma_{1-W_i}^{C_i}).$$

Specifically, $P(\tilde{\mathbf{B}}_i \mid Y_i(W_i)) = P(\tilde{B}_{i,2} \mid Y_i(0)) = \text{Lap}(\tilde{B}_{i,2} \mid Y_i(0), 1/\epsilon_{b_2})$ for $W_i = 0$ and $P(\tilde{\mathbf{B}}_i \mid Y_i(W_i)) = P(\tilde{B}_{i,1} \mid Y_i(1)) = \text{Lap}(\tilde{B}_{i,1} \mid Y_i(1), 1/\epsilon_{b_1})$ for $W_i = 1$. The privacy-aware Metropolis-within-Gibbs algorithm (Ju et al., 2022) is used for the draw of $Y_i(W_i)$.

C Simulation Details

C.1 Beta GLM

Under the data-generating processes and the re-parameterizations of the Beta regression provided in Section 5.1, we generated 1000000 samples for $Y(0)$ and $Y(1)$ to see what the

Table 4: Evaluation metrics of frequentist estimators for $N = 100, N_{sim} = 2000$.

	Coverage			Bias			MSE			Interval Width		
ϵ_{tot}	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)
0.1	94.75%	94.5%	100.0%	0.9025	−1.0975	−1.0975	0.9977	0.8231	0.7357	1.895	1.881	2.0
0.3	93.4%	94.95%	100.0%	0.9025	−0.6965	0.9025	0.9827	0.4956	0.7877	1.869	1.803	2.0
1.0	94.8%	95.45%	99.8%	−0.0535	−0.2887	0.9025	0.7787	0.084	0.7476	1.883	1.137	1.986
3.0	94.65%	94.70%	97.6%	−0.3555	0.1052	−0.844	0.1037	0.0176	0.2508	1.237	0.518	1.673
10	95.85%	95.0%	95.7%	0.1429	0.1127	−0.1147	0.0115	0.0092	0.0226	0.433	0.38	0.591

Table 5: Evaluation metrics of Bayesian estimators for $N = 100, N_{sim} = 1000$.

	Coverage			Bias			MSE			Interval Width		
ϵ_{tot}	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)
0.1	100.0%	100.0%	100.0%	−0.0961	−0.0958	−0.0966	0.00927	0.00938	0.00937	0.616	0.614	0.575
0.3	100.0%	100.0%	100.0%	−0.0958	−0.0886	−0.097	0.00922	0.00889	0.00948	0.616	0.602	0.586
1.0	100.0%	100.0%	100.0%	−0.0952	−0.0691	−0.0951	0.00932	0.00939	0.00961	0.615	0.528	0.58
3.0	99.5%	97.6%	99.5%	−0.0657	−0.0403	−0.0864	0.01055	0.00734	0.01126	0.521	0.367	0.54
10	94.4%	96.7%	94.3%	−0.0155	−0.0256	−0.0259	0.00343	0.00241	0.00676	0.232	0.198	0.304

data looks like. Figure 1 shows the distributions of each potential outcome. Also, the expectations of each potential outcome are expressed as:

$$\begin{aligned}
\mathbb{E}[Y(0)] &= \mathbb{E}_{X_1,X_2,X_3}[\mu(0)] \\
&= \mathbb{E}_{X_1,X_2,X_3}\left[\frac{\exp(1.0 - 0.8X_1 + 0.5X_2 - 2.0X_3)}{1 + \exp(1.0 - 0.8X_1 + 0.5X_2 - 2.0X_3)}\right] \\
&= 0.359613, \\
\mathbb{E}[Y(1)] &= \mathbb{E}_{X_1,X_2,X_3}[\mu(1)] \\
&= \mathbb{E}_{X_1,X_2,X_3}\left[\frac{\exp(1.5 - 0.8X_1 + 0.5X_2 - 2.0X_3)}{1 + \exp(1.5 - 0.8X_1 + 0.5X_2 - 2.0X_3)}\right] \\
&= 0.457068.
\end{aligned}$$

We refer readers to Ferrari and Cribari-Neto (2004) for further details about the Beta regression.

C.2 Additional Simulations

Table 4 – 7 display the simulation results for smaller sample sizes of $N = 100$ and $N = 1000$. All scenarios achieve roughly 95% coverage. Regarding Bias and MSE, custom scenarios demonstrate superior performance compared to the joint scenario, consistent with the observations in the main manuscript for $N = 10000$. As expected, the MSE of the frequentists estimators for $N = 1000$ is about 10 times that of $N = 10000$, and $N = 100$ is about 10 times that of $N = 1000$, which confirms the validity of the convergence rates we derived. All discussions regarding the comparison between the frequentist and Bayesian estimators in the main manuscript are applicable to the case of $N = 100, 1000$. Please refer to Section 5 in the main manuscript for a detailed discussion on this matter.

Table 6: Evaluation metrics of frequentist estimators for $N = 1000, N_{sim} = 2000$.

	Coverage			Bias			MSE			Interval Width		
ϵ_{tot}	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)
0.1	94.40%	94.55%	100.0%	−1.0975	−0.2006	0.1042	1.0006	0.4808	0.7886	1.887	1.781	1.999
0.3	95.05%	95.85%	99.55%	0.9025	0.1541	0.9025	0.9411	0.0857	0.7609	1.901	1.148	1.987
1.0	94.20%	95.20%	98.0%	−1.0975	−0.1484	0.0074	0.3919	0.0089	0.2192	1.737	0.369	1.618
3.0	95.15%	94.65%	96.3%	−0.0214	0.0245	−0.0585	0.0111	0.0018	0.0216	0.411	0.164	0.586
10	94.55%	95.05%	94.65%	−0.0082	−0.0189	0.0671	0.0012	0.0009	0.0022	0.137	0.12	0.181

Table 7: Evaluation metrics of Bayesian estimators for $N = 1000, N_{sim} = 1000$.

	Coverage			Bias			MSE			Interval Width		
ϵ_{tot}	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)	Joint	Custom (IPW)	Custom (DM)
0.1	100.0%	100.0%	100.0%	−0.0973	−0.0935	−0.0987	0.0096	0.0094	0.0099	0.45	0.448	0.462
0.3	100.0%	99.1%	100.0%	−0.0972	−0.0787	−0.0974	0.0096	0.0088	0.0099	0.455	0.411	0.463
1.0	100.0%	94.0%	100.0%	−0.0928	−0.0365	−0.0916	0.0093	0.0046	0.0099	0.445	0.259	0.433
3.0	96.1%	92.3%	95.7%	−0.0265	−0.0173	−0.0469	0.004	0.0015	0.0056	0.258	0.134	0.303
10	92.7%	95.5%	92.0%	−0.0103	−0.0074	−0.0144	0.0004	0.0003	0.0009	0.078	0.061	0.107

D Regression Adjustment

D.1 Overview

In the context of randomized experiments, causal effects τ can be identified solely using the treatment assignment and outcome variables. Also, as demonstrated in prior sections, our custom frequentist estimators achieve minimax optimality without the need for covariates. However, there is a clear rationale for incorporating covariates when deducing causal effects in randomized settings: they can enhance the efficiency of inference by leveraging pertinent individual data. This enhancement method is termed regression adjustment (Lin, 2013). Nevertheless, applying regression adjustment within LDP presents challenges. Specifically, it could incur additional privacy costs for the covariates, and these costs could escalate significantly for high-dimensional covariates. In this section, we present another type of frequentist estimator for the joint scenario, namely the OLS estimator. We explore its advantages and constraints, compared to the IPW estimator in the same scenario.

Assume that the observed covariates are privatized by the Laplace mechanism. We assume $X_{i,j} \in [0, 1]$ for $i = 1, \ldots, N$ and $j = 1, \ldots, d$ to ensure bounded ℓ_1-sensitivity. The privatized outcomes and covariates are $\tilde{X}_{i,j} = X_{i,j} + \nu^X_{i,j}$, where $\nu^X_{i,j} \sim^{i.i.d} \text{Lap}(d/\epsilon_x)$. By composition, the joint release of $(\tilde{Y}_i, \tilde{X}_{i,1}, \ldots, \tilde{X}_{i,d}, \tilde{W}_i)_{i=1}^N$ satisfies $(\epsilon_y + \epsilon_x + \epsilon_w)$-LDP.

Without privacy considerations, it is well known that the covariate adjustment can further improve the efficiency, even without assuming a correctly specified outcome model (Lin, 2013). Specifically, we propose the following plug-in OLS estimator.

$$\tilde{\tau}_{OLS} = \tilde{\alpha}_{(1)} - \tilde{\alpha}_{(0)} + \bar{X}(\tilde{\beta}_{(1)} - \tilde{\beta}_{(0)}), \tag{11}$$

where $\bar{X} = \frac{1}{N}\sum_{i=1}^N \tilde{X}_i$ and $(\tilde{\alpha}_{(w)}, \tilde{\beta}_{(w)}) = \arg\min_{\alpha,\beta} \sum_{i:\tilde{W}_i=w} (\tilde{Y}_i - \alpha - \tilde{X}_i'\beta)^2$ for $w = 0, 1$. Note that, under some regularity conditions (Lehmann and Casella, 1998, p. 440), $(\tilde{\alpha}_{(w)}, \tilde{\beta}_{(w)})$ converges to $(\tilde{\alpha}^*_{(w)}, \tilde{\beta}^*_{(w)})$, defined as

$$(\tilde{\alpha}^*_{(w)}, \tilde{\beta}^*_{(w)}) = \arg\min_{\alpha,\beta} \mathbb{E}[(\tilde{Y}_i - \alpha - \tilde{X}_i'\beta)^2 \mid \tilde{W}_i = w].$$

We investigate the potential bias of the naïve OLS estimator and propose a bias-corrected version. The following theorem states that the naïve OLS estimator (11) is an inconsistent estimator for τ, but multiplying by the same factor C_{p,ϵ_w} makes it consistent. The central limit theorem has also been developed.

Theorem D.1. *1. (Consistency) $C_{p,\epsilon_w}\tilde{\tau}_{OLS}$ is consistent for τ.*

2. (CLT) $\sqrt{N}(C_{p,\epsilon_w}\tilde{\tau}_{OLS} - \tau)$ converges in distribution to

$$\mathrm{N}\left(0, C^2_{p,\epsilon_w}\left(\frac{\mathrm{MSE}_1}{\rho_1} + \frac{\mathrm{MSE}_0}{\rho_0}\right)\right), \tag{12}$$

where $\mathrm{MSE}_w = \mathbb{E}[(\tilde{Y}_i - \tilde{\alpha}^*_{(w)} - \tilde{X}_i'\tilde{\beta}^*_{(w)})^2 \mid \tilde{W}_i = w]$ for $w = 0, 1$.

3. (Confidence Interval) The following interval is the nominal central confidence at the significance level α:

$$\left(C_{p,\epsilon_w}\tilde{\tau}_{OLS} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{OLS}}{N}}, C_{p,\epsilon_w}\tilde{\tau}_{OLS} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\Sigma}_{OLS}}{N}}\right),$$

where $\hat{\Sigma}_{OLS} = C^2_{p,\epsilon_w}\big(\frac{\widehat{\mathrm{MSE}}_1}{\rho_1} + \frac{\widehat{\mathrm{MSE}}_0}{\rho_0}\big)$ and $\widehat{\mathrm{MSE}}_w = \frac{1}{\tilde{N}_w}\sum_{i:\tilde{W}_i = w}(\tilde{Y}_i - \tilde{\alpha}_{(w)} - \tilde{X}_i\tilde{\beta}_{(w)})^2$ for $w = 0, 1$.

Proof. Consider the objective function

$$\begin{aligned}\mathcal{Q}(\alpha_{(w)}, \beta_{(w)}) &= \mathbb{E}[(\tilde{Y}_i - \alpha_{(w)} - \tilde{X}_i'\beta_{(w)})^2 \mid \tilde{W}_i = w] \\ &= \mathbb{E}[(\tilde{Y}_i - \gamma_{(w)} - (\tilde{X}_i' - \mu_{\tilde{X}})\beta_{(w)})^2 \mid \tilde{W}_i = w],\end{aligned}$$

where $\gamma_{(w)} = \alpha_{(w)} + \mu_{\tilde{X}}\beta_{(w)}$. Note that, for both $w = 0, 1$,

$$\mu_{\tilde{X}} = \mathbb{E}[\tilde{X}_i \mid \tilde{W}_i = w] = \mathbb{E}[X_i \mid \tilde{W}_i = w] = \mathbb{E}[X_i] = \mu_X.$$

The second equality follows from the independence of noise ν_i^X, and the third equality follows from the randomized assignment of W_i and the independence of the randomized response mechanism. Minimizing the right-hand side over $\gamma_{(w)}$ and $\beta_{(w)}$ leads to the same values for $\alpha_{(w)}$ and $\beta_{(w)}$ as minimizing the left-hand side over $\alpha_{(w)}$ and $\beta_{(w)}$, with the least squares estimate of $\gamma^*_{(w)} = \alpha^*_{(w)} + \mu_{\tilde{X}}\beta^*_{(w)}$.

$$\begin{aligned}&\mathcal{Q}(\gamma_{(w)}, \beta_{(w)}) \\ &= \mathbb{E}[(\tilde{Y}_i - \gamma_{(w)} - (\tilde{X}_i' - \mu_X)\beta_{(w)})^2 \mid \tilde{W}_i = w] \\ &= \mathbb{E}[(\tilde{Y}_i - \gamma_{(w)})^2 \mid \tilde{W}_i = w] + \mathbb{E}[((\tilde{X}_i' - \mu_{\tilde{X}})\beta_{(w)})^2 \mid \tilde{W}_i = w] - 2\mathbb{E}[(\tilde{Y}_i - \gamma_{(w)})(\tilde{X}_i' - \mu_{\tilde{X}})\beta_{(w)} \mid \tilde{W}_i = w] \\ &= \mathbb{E}[(\tilde{Y}_i - \gamma_{(w)})^2 \mid \tilde{W}_i = w] + \mathbb{E}[((\tilde{X}_i' - \mu_{\tilde{X}})\beta_{(w)})^2 \mid \tilde{W}_i = w] - 2\mathbb{E}[\tilde{Y}_i(\tilde{X}_i' - \mu_{\tilde{X}})\beta_{(w)} \mid \tilde{W}_i = w].\end{aligned}$$

The last two terms do not depend on $\gamma_{(w)}$. Thus, minimizing $\mathcal{Q}(\gamma_{(w)}, \beta_{(w)})$ over $\gamma_{(w)}$ is equivalent to minimizing $\mathbb{E}[(\tilde{Y}_i - \gamma_{(w)})^2 \mid \tilde{W}_i = w]$ over $\gamma_{(w)}$, which leads to the minimizer

$$\begin{aligned}
\tilde{\gamma}^*_{(1)} &= \mathbb{E}[\tilde{Y}_i|\tilde{W}_i = 1] = \mathbb{E}[Y_i|\tilde{W}_i = 1] \\
&= \sum_{w=0}^{1} \mathbb{E}[Y_i|\tilde{W}_i = 1, W_i = w] P(W_i = w \mid \tilde{W}_i = 1) \\
&= \frac{\bar{p}\bar{q}_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}} \mathbb{E}[Y_i(0)] + \frac{pq_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}} \mathbb{E}[Y_i(1)].
\end{aligned}$$

Similarly, we have

$$\tilde{\gamma}^*_{(0)} = \frac{\bar{p}q_{\epsilon_w}}{\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w}} \mathbb{E}[Y_i(0)] - \frac{p\bar{q}_{\epsilon_w}}{\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w}} \mathbb{E}[Y_i(1)].$$

Then, we have

$$\begin{aligned}
\tilde{\gamma}^*_{(1)} - \tilde{\gamma}^*_{(0)} =& \frac{(q_{\epsilon_w} - \bar{q}_{\epsilon_w})p\bar{p}}{(\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w})(pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w})} (\mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]) \\
=& \frac{(q_{\epsilon_w} - \bar{q}_{\epsilon_w})p\bar{p}}{(\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w})(pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w})} \tau \\
=& \frac{1}{C_{p,\epsilon_w}} \tau.
\end{aligned}$$

Finally, noting the fact that $\tilde{\gamma}^*_{(w)} = \tilde{\alpha}^*_{(w)} + \mu_{\tilde{X}} \tilde{\beta}^*_{(w)}$ and, under some regularity conditions, $(\tilde{\alpha}_{(w)}, \tilde{\beta}_{(w)})$ converges to $(\tilde{\alpha}^*_{(w)}, \tilde{\beta}^*_{(w)})$,

$$\tilde{\tau}_{OLS} = \tilde{\alpha}_{(1)} - \tilde{\alpha}_{(0)} + \bar{\tilde{X}}(\tilde{\beta}_{(1)} - \tilde{\beta}_{(0)}) \xrightarrow{p} \tilde{\gamma}^*_{(1)} - \tilde{\gamma}^*_{(0)} = \frac{1}{C_{p,\epsilon_w}} \tau.$$

Thus, by the continuous mapping theorem, $C_{p,\epsilon_w} \tilde{\tau}_{OLS}$ is a consistent estimator for τ.

Next, we obtain the central limit theorem. Again, it is convenient to parameterize the model using (γ_w, β_w) instead of (α_w, β_w). In terms of these parameters, the objective function for $\tilde{W}_i = w$ is

$$\sum_{i:\tilde{W}_i = w} \left(\tilde{Y}_i - \gamma - (\tilde{X}_i - \mu_{\tilde{X}})\beta\right)^2.$$

The first order conditions for the estimators $(\tilde{\gamma}_w, \tilde{\beta}_w)$ are

$$\sum_{i:\tilde{W}_i = w} \psi(\tilde{Y}_i, \tilde{X}_i, \tilde{\gamma}_w, \tilde{\beta}_w) = 0,$$

where $\psi(\cdot)$ is a two-component column vector:

$$\psi(y,x,\gamma,\beta) = \begin{pmatrix} y-\gamma-(x-\mu_{\tilde X})\beta \\ (x-\mu_{\tilde X})(y-\gamma-(x-\mu_{\tilde X})\beta) \end{pmatrix}.$$

The standard M-estimation results imply that, under standard regularity conditions, the estimator is consistent and asymptotically normally distributed:

$$\sqrt{N_w}\begin{pmatrix} \tilde\gamma_w - \tilde\gamma^*_w \\ \tilde\beta_w - \tilde\beta^*_w \end{pmatrix} \xrightarrow{D} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Gamma_w^{-1}\Delta_w(\Gamma'_w)^{-1}\right),$$

where $N_w = \sum_{i=1}^{N} \mathbb{1}(\tilde W_i = w)$ and the two components of the covariance matrix are

$$\begin{aligned} \Gamma_w &= \mathbb{E}\left[\frac{\partial}{\partial(\gamma,\beta)}\psi(\tilde Y_i, \tilde X_i, \gamma, \beta) \mid \tilde W_i = w\right]\Bigg|_{(\tilde\gamma^*_w, \tilde\beta^*_w)} \\ &= \mathbb{E}\left[\begin{pmatrix} -1 & -(\tilde X_i - \mu_{\tilde X}) \\ -(\tilde X_i - \mu_{\tilde X})' & -(\tilde X_i - \mu_{\tilde X})'(\tilde X_i - \mu_{\tilde X}) \end{pmatrix} \mid \tilde W_i = w\right] \\ &= \mathbb{E}\left[\begin{pmatrix} -1 & 0 \\ 0 & -\mathbb{E}[(\tilde X_i - \mu_{\tilde X})'(\tilde X_i - \mu_{\tilde X})] \end{pmatrix} \mid \tilde W_i = w\right], \end{aligned}$$

and

$$\begin{aligned} \Delta_w &= \mathbb{E}\left[\psi(\tilde Y_i, \tilde X_i, \tilde\gamma^*_w, \tilde\beta^*_w) \cdot \psi(\tilde Y_i, \tilde X_i, \tilde\gamma^*_w, \tilde\beta^*_w)' \mid \tilde W_i = w\right] \\ &= \mathbb{E}\left[(\tilde Y_i - \tilde\gamma^*_w - (\tilde X_i - \mu_{\tilde X})\tilde\beta^*_w)^2 \cdot \begin{pmatrix} -1 \\ (\tilde X_i - \mu_{\tilde X})' \end{pmatrix} \begin{pmatrix} -1 \\ (\tilde X_i - \mu_{\tilde X})' \end{pmatrix}' \mid \tilde W_i = w\right]. \end{aligned}$$

The variance of $\tilde\gamma_w$ is the $(1,1)$ element of the covariance matrix. Because Γ_w is block diagonal, the $(1,1)$ element is equal to

$$\begin{aligned} \mathrm{MSE}_w &= \mathbb{E}[(\tilde Y_i - \tilde\gamma^*_w - (\tilde X_i - \mu_{\tilde X})\tilde\beta^*_w)^2 \mid \tilde W_i = w] \\ &= \mathbb{E}[(\tilde Y_i - \tilde\alpha^*_w - \tilde X_i'\tilde\beta^*_w)^2 \mid \tilde W_i = w]. \end{aligned}$$

Therefore, we have

$$\sqrt{N_w}\begin{pmatrix} \tilde\gamma_w - \tilde\gamma^*_w \\ \tilde\beta_w - \tilde\beta^*_w \end{pmatrix} \xrightarrow{D} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathrm{MSE}_w \begin{pmatrix} 1 & 0 \\ 0 & \left(\mathbb{E}[(\tilde X_i - \mu_{\tilde X})'(\tilde X_i - \mu_{\tilde X})]\right)^{-1} \end{pmatrix}\right),$$

which implies

$$\sqrt{N}(\tilde\gamma_{(w)} - \tilde\gamma^*_{(w)}) \xrightarrow{D} \mathrm{N}\left(0, \frac{\mathrm{MSE}_w}{P(\tilde W_i = w)}\right). \tag{13}$$

As shown before, $\tau = C_{p,\epsilon_w}(\tilde\gamma^*_1 - \tilde\gamma^*_0)$. Also, $C_{p,\epsilon_w}\tilde\tau_{OLS} = C_{p,\epsilon_w}(\tilde\gamma_1 - \tilde\gamma_0) = C_{p,\epsilon_w}\{\tilde\alpha_1 - \tilde\alpha_0 + \bar{\tilde X}(\tilde\beta_1 - \tilde\beta_0)\}$ is the consistent estimator for τ. Noting that $\tilde\beta_1$, $\tilde\beta_0$, $\tilde\gamma_1$ and $\tilde\gamma_0$ are all

Table 8: Evaluation metrics for the naïve and OLS estimators ($N = 1000, N_{sim} = 2000$) under the joint scenario. N_{sim} denotes the number of simulations. ϵ_{tot} denotes the total privacy budget, $\epsilon_{\text{tot}} = \epsilon_x + \epsilon_y + \epsilon_w$.

		Coverage		Bias		MSE		Interval Width	
ϵ_{tot}	$(\epsilon_x, \epsilon_y, \epsilon_w)$	Naïve	OLS	Naïve	OLS	Naïve	OLS	Naïve	OLS
3	$(1, 1, 1)$	95.3%	95.7%	−0.00266	−0.00329	0.0405	0.0371	0.798	0.770
9	$(3, 3, 3)$	95.4%	96.4%	−0.00105	−0.000422	0.00208	0.00126	0.181	0.142
30	$(10, 10, 10)$	95.0%	96.8%	−0.000547	−0.000282	0.000906	0.000177	0.120	0.058
0.3	$(0.1, 0.1, 0.1)$	95.5%	95.5%	−0.129	−0.128	0.989	0.984	1.909	1.910
3	$(2, 0.5, 0.5)$	94.5%	94.5%	−0.00703	−0.00837	0.378	0.375	1.748	1.749
3	$(0.5, 2, 0.5)$	95.2%	95.4%	−0.00576	−0.00406	0.0484	0.0373	0.857	0.754
3	$(0.5, 0.5, 2)$	95.4%	95.0%	−0.00238	−0.00263	0.0575	0.0565	0.929	0.923
3	$(0.5, 1.25, 1.25)$	94.6%	94.5%	0.00480	0.00276	0.0210	0.0187	0.547	0.518
3	$(1.25, 0.5, 1.25)$	95.3%	95.2%	−0.00101	−0.00246	0.103	0.101	1.232	1.225
3	$(1.25, 1.25, 0.5)$	94.6%	95.7%	0.00137	0.00150	0.102	0.0889	1.195	1.144

asymptotically independent, the asymptotic distribution of $\tilde{\tau}_{OLS}$ is expressed as

$$\sqrt{N}(C_{p,\epsilon_w}\hat{\tau}_{OLS} - \tau) \xrightarrow{D} \mathrm{N}\left(0, C^2_{p,\epsilon_w}\left(\frac{\mathrm{MSE}_1}{\rho_1} + \frac{\mathrm{MSE}_0}{\rho_0}\right)\right).$$

□

D.2 Simulation setups for Regression Adjustment

We empirically evaluate the frequentist properties of the OLS estimator developed in Section D. We consider the joint privacy mechanism in Section 3.2 and use the same data-generating mechanisms in Section 5. We release $X_{i,d}$ after applying the Laplace mechanism. Specifically, the generated covariates satisfy the following sensitivity: $\Delta_X = 3$. Accordingly, we add the Laplace noise $\mathrm{Lap}(3/\epsilon_y)$ to $X_{i,k}$ for $k = 1, 2, 3$. Then, we obtain the private data $\tilde{X}_{i,k}, \tilde{Y}_i, \tilde{W}_i$. By composition, this privacy mechanism guarantees that $(\tilde{Y}_i, \tilde{W}_i)$ satisfies $(\epsilon_y + \epsilon_w)$-DP and $(\tilde{X}_{i,k}, \tilde{Y}_i, \tilde{W}_i)$ satisfies $(\epsilon_x + \epsilon_y + \epsilon_w)$-DP.

D.3 Results

Tables 8 and 9 present the performance evaluation of the naïve and OLS estimators for $N = 1000, 10000$ with various privacy budgets for ϵ_x, ϵ_y and ϵ_w. We let $\epsilon_{tot} = \epsilon_x + \epsilon_y + \epsilon_w$. Both estimators achieve about 95% coverage for $N = 1000, 10000$ as expected. For bias and MSE, we observe smaller bias and MSE for larger privacy budgets. For the same levels of privacy budgets, both bias and MSE improve when N increases, which empirically supports our consistency and asymptotically unbiased properties of the estimators.

When we have a tight privacy budget of $(\epsilon_x, \epsilon_y, \epsilon_w) = (0.1, 0.1, 0.1)$, the length of the confidence interval of the frequentist estimators is nearly 2, which is almost non-informative about the estimand. When N increases, the interval length gets smaller and becomes informative enough for some allocations, e.g., $(\epsilon_x, \epsilon_y, \epsilon_w) = (1.25, 0.5, 1.25)$. However, with strict budget constraints and a small sample size, the analysis results may tell us little about the estimands, even though their consistency and confidence intervals are statistically valid. This is an inevitable trade-off between privacy protection and the accuracy of the results.

Table 9: Evaluation metrics for the naïve and OLS estimators ($N = 10000, N_{sim} = 2000$) under the joint scenario.

		Coverage		Bias		MSE		Interval Width	
ϵ_{tot}	$(\epsilon_x, \epsilon_y, \epsilon_w)$	Naïve	OLS	Naïve	OLS	Naïve	OLS	Naïve	OLS
3	$(1, 1, 1)$	95.4%	95.2%	−0.00174	−0.00196	0.00407	0.00376	0.252	0.243
9	$(3, 3, 3)$	94.7%	94.7%	−0.000154	−0.000149	0.000216	0.000136	0.0573	0.0454
30	$(10, 10, 10)$	94.6%	96.3%	0.000213	−0.0000316	0.0000962	0.0000184	0.0380	0.0183
0.3	$(0.1, 0.1, 0.1)$	94.4%	94.3%	−0.104	−0.101	0.919	0.921	1.883	1.885
3	$(2, 0.5, 0.5)$	94.9%	95.1%	−0.00380	−0.00358	0.0535	0.0520	0.915	0.906
3	$(0.5, 2, 0.5)$	95.7%	95.7%	0.00112	0.000358	0.00466	0.00356	0.271	0.237
3	$(0.5, 0.5, 2)$	95.9%	95.9%	0.000703	0.000989	0.00524	0.00512	0.295	0.292
3	$(0.5, 1.25, 1.25)$	95.9%	95.9%	0.00133	0.00124	0.00187	0.00170	0.173	0.163
3	$(1.25, 0.5, 1.25)$	95.1%	95.0%	−0.000968	−0.000691	0.0106	0.0104	0.405	0.401
3	$(1.25, 1.25, 0.5)$	95.4%	95.4%	0.00247	0.00279	0.00957	0.00848	0.391	0.369

D.4 Discussions

In the simulations, we consider different divisions of the same overall privacy budget, $\epsilon_{tot} = 3$, which suggests an allocation strategy of the budget. Among all the budget allocations with $\epsilon_{tot} = 3$, we see that $(\epsilon_x, \epsilon_y, \epsilon_w) = (0.5, 1.25, 1.25)$ achieves the lowest MSE for both naïve and OLS estimators. Thus, it seems reasonable to assign a strict budget to X, and larger budgets to Y and W. We also see that for most allocations with budgets $\epsilon_{tot} \leq 3$, there is minimal gain in MSE for the OLS over the naïve estimator. However, for $(\epsilon_x, \epsilon_y, \epsilon_w) = (10, 10, 10), (3, 3, 3), (0.5, 2, 0.5)$, we see that the OLS estimator does significantly outperform the naïve estimator in terms of MSE. This result follows from the fact that the regression adjustment technique in randomized experiments (Freedman, 2008; Lin, 2013) helps reduce the variance of the OLS estimator, leading to better MSE. Intuitively, the regression adjustment works for $(\epsilon_x, \epsilon_y, \epsilon_w) = (10, 10, 10)$ because the privatized data contains smaller noise, and $\tilde{X}$ still contains some information to explain $\tilde{Y}$. When the total budget is smaller ($\epsilon_{\text{tot}} \leq 3$), however, the gain is limited.

We here further discuss some limitations to the gains in precision of the estimator for the PATE from including covariates from theoretical perspectives. In large samples, including covariates in the regression function under usual randomized experiments will not lower the precision (Imbens and Rubin, 2015). However, DP mechanisms under randomization pose unique challenges. First, MSE_w in Theorem D.1 can be written as follows:

$$\text{MSE}_w = \mathbb{V}\text{ar}[Y_i|\tilde{W}_i = w] + \mathbb{E}[Y_i|\tilde{W}_i = w]^2 + \frac{1}{\epsilon_y^2} - \mathbb{E}[\tilde{Y}_i'\tilde{X}_i(\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'\tilde{Y}_i|\tilde{W}_i = w]. \quad (14)$$

The last term, $\mathbb{E}[\tilde{Y}_i'\tilde{X}_i(\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'\tilde{Y}_i|\tilde{W}_i = w]$, is effectively the gain in precision from including covariates. This term implies that the gain is zero when $\tilde{X}_i$ and $\tilde{Y}_i$ are orthogonal, but is always positive otherwise. As adding large independent noise to X_i and Y_i makes the privatized observations less correlated, the gain becomes negligible when ϵ_x and ϵ_y are small. We also note that the first two terms in (14) are bounded due to the sensitivity of Y; however, the last two terms are unbounded, making them the dominant precision factors, especially when ϵ_x and ϵ_y are small. Therefore, the gain from adding covariates in inference is actually limited in our LDP scenarios.