Locally Private Causal Inference

Yuki Ohnishi
Department of Statistics
Purdue University
West Lafayette, IN 47907
yohnishi@purdue.edu

Jordan Awan
Department of Statistics
Purdue University
West Lafayette, IN 47907
jawan@purdue.edu

Abstract

Local differential privacy (LDP) is a differential privacy (DP) paradigm in which individuals first apply a DP mechanism to their data (often by adding noise) before transmitting the result to a curator. In this article, we develop methodologies to infer causal effects from locally privatized data under the Rubin Causal Model framework. First, we present frequentist estimators under various privacy scenarios with their variance estimators and plug-in confidence intervals. We show that using a plug-in 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.

1 Introduction

Causal inference is a fundamental consideration across a wide range of domains in science, technology, engineering, and medicine (Imbens & Rubin, 2015). Researchers study randomized experiments or observational studies to unveil the causal effects of treatment assignment in an unbiased manner with valid uncertainty quantification. 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 the 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 for more stringent privacy protection (e.g., Erlingsson et al., 2014; Apple, 2017).

In this paper, we consider three distinct local privacy scenarios. The first scenario assumes that all accessible variables require privacy protection. In the second and third scenarios, we are allowed to select the variable we privatize. We then offer causal inferential methodologies to analyze such privatized data. Our main contributions are as follows:

Preprint. Under review.

- First, we propose a “naïve” inverse probability weighting (IPW) estimator. We then compute a bias of the IPW estimator and propose a bias correction technique.
- We propose efficient frequentist estimators that achieve the optimal convergence rate under custom scenarios where we are allowed to select the variable we privatize.
- We also compute the asymptotic variance, and construct asymptotic plugin nominal confidence intervals for all frequentist estimators. We also discuss their optimality under distinct privacy scenarios.
- We develop a Bayesian nonparametric methodology. To implement the methodology, we develop a data augmentation Gibbs sampler, which can be applied to all scenarios that are considered in the frequentist analyses.
- We present simulation studies to evaluate the frequentist and Bayesian methodologies at various privacy budgets to understand how they perform in practice. .

This paper is organized as follows. In Section 2 we review related work. Section 3 presents the preliminaries for the Rubin Causal Model and LDP. In Section 4, we develop statistically valid frequentist approaches to inferring the causal effects of interest. Section 5 presents a Bayesian methodology for performing valid causal inference with the privatized data. Section 6 provides simulation studies for validating our methodologies developed in the previous sections, and Section 7 concludes with some final discussion. Finally, all proofs are left to the Supplementary Materials.

2 Related Work

Despite the fact that DP is a rapidly growing field, statistically valid causal inferential methodologies for differentially private data are limited. The following work uses LDP for its DP mechanism. Agarwal & Singh (2021) proposed an end-to-end procedure for data cleaning, estimation, and inference with data cleaning-adjusted confidence intervals under the local differential privacy mechanism. They consider a general case where only the covariates are corrupted, e.g., privatized, missing, or measured with errors.

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 & 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 differential privacy. Niu et al. (2022) introduced a general meta-algorithm for privately estimating conditional average treatment effects.

Finally, Imai & Yamamoto (2010) studied the nonparametric identification of the average treatment effect when a treatment variable is measured with errors. Their work did not consider differential privacy and solely addressed a case where only treatment variables have measurement errors.

In non-causal domains, Evans & 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.

3 Preliminaries

3.1 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. We present an illustration of the RCM in Appendix for clarity. 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 & 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 experimental unit i has an observed outcome Y_i and treatment assignment W_i for unit i respectively. We consider a binary treatment $W_i = w \in \{0, 1\}$ with $p(W_i = 1) = p$ which is assumed to be known by the experimental design, and let $Y_i(0)$ and $Y_i(1)$ denote potential outcomes for $W_i = w \in \{0, 1\}$. In this article, we consider the N units as a random sample from a large superpopulation, 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 PATE by the estimators derived in this manuscript.

Assumption 3.1.

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: $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$.

3.2 Differential Privacy

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

Definition 3.2 (Local Differential Privacy)**.** An algorithm $\mathcal{M}$ is said to be ϵ-locally differentially private (ϵ-LDP) if for any two data points $x, x' \in D$, 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. 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 a 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: $D \to \mathbb{R}^k$ is $\Delta_f = \sup_{x,y \in D} ||f(x) - f(y)||_1$.

Proposition 3.3 (Laplace Mechanism)**.** *Let $f : D \to \mathbb{R}^k$. The Laplace mechanism is defined as*

$$M(D) = f(D) + (\nu_1, ..., \nu_k)^\top,$$

where the ν_i are independent Laplace random variables, $\nu_i \sim \mathrm{Lap}(0, \Delta f/\epsilon)$, where the density of the Laplace distribution, $\mathrm{Lap}(\mu, b)$, *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 3.4 (Randomized Response Mechanism)**.** *Let $Z_i \in \{0, 1\}$ be a binary variable. The randomized response mechanism is*

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

which satisfies ϵ-LDP.

4 Frequentist Approach

In this section, we discuss frequentist estimators for τ under several scenarios where variables are privatized in a different manner.

4.1 Minimax Risk Lower Bound for PATE Estimation

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 4.1, 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 $i = 1, \dots, N$, 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. 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 4.1. *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 4.1 implies that the optimal estimator of the PATE estimation problem also has the minimax lower bound $O((N\epsilon^2)^{-1})$.

4.2 Joint scenario, known p

We first consider a scenario where all variables are jointly 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 4.2. *Under Assumption 3.1, 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 4.3, 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 4.3. 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 normal distribution and the confidence interval construction are in Appendix D. Setting $\epsilon_y = \epsilon_w = \epsilon/2$ gives MSE of $O((N\epsilon^4)^{-1})$, which does not match the optimal rate (1). In the following sections, we see that when we use a customized privacy mechanism, rather than a naïve joint privatization, we are able to match the minimax lower bound.

4.3 Custom scenario, 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 4.4. 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 Appendix E. We see in Theorem 4.4 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 4.2.

4.4 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 3.1 remains valid. An example of this setting is A/B testing, where marketers assign treatments with an undisclosed probability without looking at 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

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

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 4.5. *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 Appendix F. 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.

4.5 Remarks

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. In contrast, access to the complete dataset is unavailable in the two custom scenarios, but they attain the optimal rate of the locally private PATE estimation.

When the sample size is small, and 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.

In Appendix, we introduce another class of frequentist estimators under the joint scenario, the difference-in-means (DM) and OLS estimators. We discuss the benefits and limitations of the OLS estimator over the IPW estimator under the joint scenario.

5 Bayesian Approach

Following the Bayesian paradigm of Imbens & Rubin (2015), we consider deriving the posterior distributions of the causal estimands (Forastiere et al., 2016; Ohnishi & Sabbaghi, 2022a). The key idea is the data augmentation (Tanner & Wong, 1987) to obtain the posterior distribution of the causal estimands by imputing in turn the missing variables. The idea of the data augmentation for estimating causal effects is outlined in Imbens & Rubin (1997), but our difficulties 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. In what follows, we focus on the joint scenario discussed in Section 4.2 to show the outline of our algorithm, but it can easily be extended to the custom scenarios, as explained in Appendix.

Under the super-population perspective, these quantities 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 unobserved 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

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

which follows from the conditional independence of potential outcomes and $\tilde{W}_i$ given W_i (Lemma B.1 in Appendix) 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 $p(Y_i(0), Y_i(1) \mid \boldsymbol{\theta}_Y)$. Under this modeling strategy, our Bayesian approach is a valid inference for PATE. Note that PATE is a function of the parameters $\boldsymbol{\theta}_Y$, which governs the potential outcomes. Thus, it suffices to obtain the posterior draws of the posterior of the $\boldsymbol{\theta}_Y$ for the posterior draws of PATE.

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

5.1 Algorithm Outlines

Equation (5) motivates the Gibbs sampling procedures to obtain the draws from the posterior distribution of $\boldsymbol{\theta}_Y$. This section describes the key steps of the Gibbs sampler. The technical details of each step are provided in the Supplementary Materials. The key steps of the algorithm proceed 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 $r_w = p(\tilde{Y}_i | Y_i(w)) p(\tilde{W}_i | W_i = w) p(W_i = w)$ for $w = 0, 1$.
2. Given $\boldsymbol{\theta}_Y$ and W_i, draw each $Y_i(0)$ and $Y_i(1)$ according to:
$$\begin{aligned} p(Y_i(W_i)|-) &\propto p(Y_i(W_i) \mid W_i, \boldsymbol{\theta}_Y) p(\tilde{Y}_i \mid Y_i(W_i)) \\ p(Y_i(1 - W_i)|-) &\propto p(Y_i(1 - W_i) \mid W_i, \boldsymbol{\theta}_Y). \end{aligned}$$
3. Update model parameters via the blocked Gibbs sampler and calculate the estimands.

Each step is derived from the corresponding components of (5). 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 missing potential outcomes $Y_i(1 - W_i)$ are just generated from the outcome model $p(Y_i(1-W_i) \mid W_i, \boldsymbol{\theta}_Y)$, whereas the posterior distribution of $Y_i(W_i)$ cannot be obtained in a closed form as it is weighted by the privacy mechanism $p(\tilde{Y}_i \mid Y_i(W_i))$. 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 $p(Y_i(0), Y_i(1)|\theta_Y)$ at the current value of θ_Y, rather than specifying a custom proposal distribution and step size for the Metropolis-Hastings step. Finally, Step 3 updates all the parameters of the DPM that govern the potential outcomes, using standard techniques; see Section J of the Supplementary Materials for full details of the algorithm and the extension of the algorithm to the custom scenarios, which requires slight modifications of Step 1 and 2.

6 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.

6.1 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 & 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. The details are provided in the Supplementary Materials.

6.2 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 4.2, 4.3, and 4.4 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 4.2, 4.3, and 4.4, 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

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 4.3 and 4.4 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	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 for Bayes estimators ($N = 10000, N_{sim} = 200$).

	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	96.0%	94.0%	95.0%	−0.0942	−0.0809	−0.0985	0.0098	0.0087	0.0107	0.356	0.318	0.346
0.3	96.0%	95.0%	97.5%	−0.0982	−0.033	−0.0839	0.0107	0.0032	0.0085	0.349	0.218	0.344
1.0	95.0%	93.5%	90.5%	−0.0705	−0.0094	−0.0582	0.0079	0.0006	0.0063	0.325	0.096	0.267
3.0	86.0%	93.5%	91.0%	−0.0095	−0.0063	−0.0069	0.0008	0.0002	0.0011	0.088	0.044	0.114
10	95.0%	94.0%	89.5%	−0.0022	−0.0022	−0.0042	0.0	0.0	0.0001	0.026	0.022	0.035

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. One limitation of the Bayesian approach is that the coverage of the posterior interval is not necessarily well-calibrated. We observe a less-calibrated coverage probability, 86%, for $\epsilon_{\text{tot}} = 3.0$ under the joint scenario and 89.5% for $\epsilon_{\text{tot}} = 10$ under the custom (DM) scenario. This is because the posterior credible intervals generally have no guarantee to achieve the nominal coverage probability. We will leave this issue for future investigation. We provide additional simulation studies to evaluate the performance of our OLS estimator, which is derived in Appendix.

7 Concluding Remarks

In this article we proposed causal inferential methodologies to analyze differential private data under the Rubin Causal Model. We considered three different local privacy scenarios that have practical relevance. First, we presented frequentist estimators under the different privacy scenarios with their variance estimators, plug-in confidence intervals and convergence rate. We showed that using a plug-in estimator results in poor MSE compared to the minimax lower bounds. In contrast, we show that by using a customized privacy mechanism, we can achieve the lower bound and obtain minimax optimal inference. Finally we presented a Bayesian methodology and its sampling algorithm, as an alternative to the frequentist methodologies. We mainly discussed the Laplace and randomized response mechanism for privacy mechanisms for simplicity. However, 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.

A direction for future research is to develop an analytical framework for unbounded X and Y. Our framework is restricted for bounded X and Y due to considerations of the sensitivity of differential privacy mechanisms. The finite-sample performance of our estimators may be improved by more carefully choosing the noise adding mechanisms; one may investigate using truncated-uniform-Laplace (Tulap) (Awan & Slavković, 2018), K-norm mechanisms (Hardt & Talwar, 2010; Awan & Slavković, 2021), or the minimax optimal noise mechanism for multivariate mean estimation (Duchi et al., 2018). Another direction of future work would be to develop minimax optimal LDP estimators of PATE in observational studies.

References

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

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

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

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

Cribari-Neto, F., Garcia, N. L., and Vasconcellos, K. L. P. A note on inverse moments of binomial variates. *Brazilian Review of Econometrics*, 20(2), 2000.

D'Orazio, V., Honaker, J., and King, G. Differential privacy for social science inference. *SSRN Electronic Journal*, 01 2015. doi: 10.2139/ssrn.2676160.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018. doi: 10.1080/01621459.2017.1389735.

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

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

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

Ferguson, T. S. Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2(4): 615 – 629, 1974. doi: 10.1214/aos/1176342752.

Ferrari, S. and Cribari-Neto, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815, 2004. doi: 10.1080/0266476042000214501.

Forastiere, L., Mealli, F., and VanderWeele, T. J. 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, 2016. ISSN 1537274X. doi: 10.1080/01621459.2015.1125788.

Freedman, D. A. On regression adjustments in experiments with several treatments. *The Annals of Applied Statistics*, 2(1):176 – 196, 2008. doi: 10.1214/07-AOAS143.

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

Imai, K. and Yamamoto, T. Causal inference with differential measurement error: Nonparametric identification and sensitivity analysis. *American Journal of Political Science*, 54(2):543–560, 2010. ISSN 00925853, 15405907.

Imbens, G. W. and Rubin, D. B. Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25(1):305–327, 1997. ISSN 00905364.

Imbens, G. W. and Rubin, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015. doi: 10.1017/CBO9781139025751.

Ishwaran, H. and James, L. F. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001. doi: 10.1198/016214501750332758.

Ishwaran, H. and Zarepour, M. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87(2):371–390, 2000. ISSN 00063444.

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

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

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

Lehmann, E. L. and Casella, G. *Theory of Point Estimation*. Springer-Verlag, New York, NY, USA, second edition, 1998.

Lei, L. and Ding, P. Regression adjustment in completely randomized experiments with a diverging number of covariates. *Biometrika*, 108(4):815–828, 12 2020. ISSN 0006-3444. doi: 10.1093/biomet/asaa103.

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

Ohnishi, Y. and Sabbaghi, A. A Bayesian analysis of two-stage randomized experiments in the presence of interference, treatment nonadherence, and missing outcomes. *Bayesian Analysis*, pp. 1 – 30, 2022a. doi: 10.1214/22-BA1347.

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

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

Schwartz, S. L., Li, F., and Mealli, F. A Bayesian semiparametric approach to intermediate variables in causal inference. *Journal of the American Statistical Association*, 106:1331–1344, 12 2011. ISSN 01621459. doi: 10.1198/jasa.2011.ap10425.

Tanner, M. A. and Wong, W. H. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540, 1987. ISSN 01621459.

A An Example of the Rubin Causal Model

For those who are from non-causal domains, we present a simple example of the Rubin Causal Model. We consider an evaluation of a new drug in a clinical trial. Let's say we are interested in the effect of aspirin on headaches and conduct a usual treatment/placebo randomized experiment. W_i indicates the treatment assignment for unit i, that is, $W_i = 1$ if unit i is assigned to the aspirin, and $W_i = 0$ if unit i is assigned to the placebo. X_i denotes the covariates of the unit, such as age, gender, and race. They are unit-specific background attributes and are a priori known to be unaffected by the treatment assignment. Y_i generally denotes the outcomes of interest that are observed in the experiment, and $Y_i(0)$ and $Y_i(1)$ denote the "potential outcomes" for the unit. The potential outcomes are the outcomes that are potentially observed when the unit is assigned to a certain treatment condition. $Y_i(W_i)$ is the observed potential outcome, whereas $Y_i(1 - W_i)$ is the missing potential outcome that would have been observed if unit i had been assigned to the other condition of treatment. Therefore, the observed outcome for the unit would be $Y_i = Y_i(1)$ if they were assigned to the treatment (aspirin) ($W_i = 1$), and $Y_i = Y_i(0)$ if they were assigned to the placebo condition. The value of Y can be continuous or discrete, depending on the context. For example, $Y_i(w) = 1$ if they do not have a headache and $Y_i(w) = 0$ if they do have a headache when they are assigned to $W_i = w \in 0, 1$. Finally, τ denotes the causal quantities of our ultimate interest, which we call causal estimands. In the manuscript, τ is the population average difference between two potential outcomes.

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

Lemma B.1. *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 important because it plays a crucial role in proving the upcoming theorems.

C Inverse Moments of Binomial variates (Cribari-Neto et al., 2000)

Lemma C.1. *Let X be a binomial random variable such that $X \sim \mathrm{Binomial(N, p)}$. Then,*

$$\mathbb{E}[(1 + X)^{-\alpha}] = O((Np)^{-\alpha}),$$

for all $\alpha \in \mathbb{R}$.

D Details of Theorem 4.3

We provide the following central limit theorem.

Theorem D.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*

$$\mathrm{N}\left(0, C^2_{p,\epsilon_w}\left(\frac{1}{\rho_1}V_1 + \frac{1}{\rho_0}V_0 + \frac{\rho_0}{\rho_1}E_1^2 + \frac{\rho_1}{\rho_0}E_0^2 + 2E_0E_1\right)\right), \tag{6}$$

where, for $w = 0, 1$, $V_w = \mathbb{V}\mathrm{ar}(\tilde{Y}_i | \tilde{W}_i = w)$

$$= p(W_i = 0|\tilde{W}_i = w)\mathbb{V}\mathrm{ar}[Y_i(0)] + p(W_i = 1|\tilde{W}_i = w)\mathbb{V}\mathrm{ar}[Y_i(1)] + p(W_i = 0|\tilde{W}_i = w)p(W_i = 1|\tilde{W}_i = w)\tau^2 + \frac{2}{\epsilon_y^2},$$

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

We now turn to estimating the asymptotic variance of $C_{p,\epsilon_w} \tilde{\tau}_{naive}$ in (6). 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 D.2. *$\hat{V}_w$ and $\hat{E}_w$ are consistent for V_w and E_w respectively. Also, we have*

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

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 $\hat{\Sigma}_{naive} = C^2_{p,\epsilon_w}(\frac{1}{\rho_1}\hat{V}_1 + \frac{1}{\rho_0}\hat{V}_0 + \frac{\rho_0}{\rho_1}\hat{E}_1^2 + \frac{\rho_1}{\rho_0}\hat{E}_0^2 + 2\hat{E}_0\hat{E}_1)$, which is a consistent estimator for the asymptotic variance in (6).

Finally, we discuss the optimality of the naïve estimator.

Corollary D.3 (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 $O((N\epsilon^4)^{-1})$. On the other hand, the minimax lower bound for the mean estimation is $O((N\epsilon^2)^{-1})$ (Duchi et al., 2018), which corresponds to a lower bound for PATE in the case that it is known that $\mathbb{E}[Y_i(0)] = 0$, and we have $p = 1$.

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.

E Details of Theorem 4.4

By the standard central limit theorem, we have

$$\sqrt{N}(\tilde{\tau}_{IPW} - \tau) \overset{D}{\to} \mathrm{N}\left(0, \frac{\mu_1^2 + \sigma_1^2}{p} + \frac{\mu_0^2 + \sigma_0^2}{1-p} - \tau^2 - \mu_0\mu_1 + \frac{2\Delta_A}{\epsilon_a^2}\right), \tag{7}$$

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 (7).

F Details of Theorem 4.5

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

Theorem F.1. *$\tilde{\tau}_{DM}$ is consistent for τ and $\sqrt{N}(\tilde{\tau}_{DM} - \tau)$ converges in distribution to*

$$\mathrm{N}\left(0, 4\mu_0\mu_1 + \frac{\sigma_0^2}{1-p} + \frac{\sigma_1^2}{p} + \frac{2}{\epsilon_{b_1}^2}\left(\frac{\mu_0}{1-p} + \frac{\mu_1}{p}\right)^2 + \frac{2}{p^2\epsilon_{b_2}^2} + \frac{2}{(1-p)^2\epsilon_{b_3}^2}\right). \tag{8}$$

By Theorem F.1, 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 (8).

G Another class of frequentist estimators

G.1 Difference-in-means Estimator

We consider the difference-in-means (DM) estimator $\tilde{\tau}_{DM}$. A naïve DM estimator is defined by plugging in privatized observations for the usual DM estimator.

$$\tilde{\tau}_{DM} = \frac{1}{\tilde{N}_1}\sum_{i=1}^{N}\tilde{W}_i\tilde{Y}_i - \frac{1}{\tilde{N}_0}\sum_{i=1}^{N}(1-\tilde{W}_i)\tilde{Y}_i, \tag{9}$$

for $0 < \tilde{N}_w < N$ where $\tilde{N}_w = \sum_{i=1}^{N}\mathbb{1}(\tilde{W}_i = w)$. For completeness, we define $\tilde{\tau}_{DM} = 0$ for $\tilde{N}_0 = 0$ or $\tilde{N}_1 = 0$, so that the estimator is well-defined. We do not need to worry too much about this special case since it is less likely to occur for a sufficiently large number of samples under well-designed experiments, i.e., $N > 100, p = 0.5$.

The following lemma quantifies the bias of the estimator (9).

Lemma G.1. *Under Assumption 3.1, the estimator* (9) *is biased for* τ. *The bias is*

$$\mathrm{Bias}(\tilde{\tau}_{DM}) = \left(\frac{1}{C_{p,\epsilon_w}} - \rho_1^N - \rho_0^N - 1\right)\tau,$$

where $C_{p,\epsilon_w} = \frac{\rho_0\rho_1}{p(1-p)(2q_{\epsilon_w}-1)}$ *with* $\rho_w = p(\tilde{W}_i = w)$ *for* $w = 0, 1,$ *and* $q_{\epsilon_w} = \exp(\epsilon)/(1+\exp(\epsilon))$.

Note that ρ_w is a known marginal probability expressed by p and q_{ϵ_w}. This result implies that the bias only depends on the privacy budget of W. The following theorem is an immediate consequence of Lemma G.1.

Lemma G.2. $C_{p,\epsilon_w}\tilde{\tau}_{DM}$ *is a consistent and asymptotically unbiased estimator for* τ.

The asymptotic unbiasedness of $C_{p,\epsilon_w}\tilde{\tau}_{DM}$ is immediate from Lemma G.1 because all terms except for C_{p,ϵ_w} are asymptotically zero. We next focus on deriving the variance of $C_{p,\epsilon_w}\tilde{\tau}_{DM}$.

Lemma G.3.

$$\mathbb{V}\mathrm{ar}[C_{p,\epsilon_w}\tilde{\tau}_{DM}] = C_{p,\epsilon_w}^2\sum_{k=1}^{N-1}\binom{N}{k}\rho_1^k\rho_0^{N-k}\left[\frac{V_1}{k} + \frac{V_0}{N-k} + \frac{2N}{k(N-k)\epsilon_y^2}\right] + O\left((pq_{\epsilon_w})^N\right) = O\left(\frac{1}{N\epsilon_y^2\epsilon_w^2}\right),$$

where

$$V_w = p(W_i = 0|\tilde{W}_i = w)\mathbb{V}\mathrm{ar}[Y_i(0)] + p(W_i = 1|\tilde{W}_i = w)\mathbb{V}\mathrm{ar}[Y_i(1)] + p(W_i = 0|\tilde{W}_i = w)p(W_i = 1|\tilde{W}_i = w)\tau^2.$$

Note that $O((pq_{\epsilon_w})^N)$ vanishes rapidly as $N \to \infty$. The last line is implied by the lemma of Cribari-Neto et al. (2000).

We now turn to estimating the variance of $C_{p,\epsilon_w}\tilde{\tau}_{DM}$.

Lemma G.4. *$\hat{\mathbb{V}}_{DM}$ is an asymptotically unbiased estimator for $\mathbb{V}\mathrm{ar}[C_{p,\epsilon_w}\tilde{\tau}_{DM}]$.*

$$\hat{\mathbb{V}}_{DM} = C^2_{p,\epsilon_w}\left\{\hat{s}^2_1\left(N\rho_1\mathbb{E}\left[\frac{1}{(B_1+1)^2}\right] - \frac{\rho_1^N}{N}\right) + \hat{s}^2_0\left(N\rho_0\mathbb{E}\left[\frac{1}{(B_0+1)^2}\right] - \frac{\rho_0^N}{N}\right)\right\}$$

where $\hat{s}^2_w = \frac{1}{\tilde{N}_w - 1}\sum_{i:\tilde{W}_i=w}(\tilde{Y}_i - \bar{Y}_w)^2$ with $\bar{Y}_w = \frac{1}{\tilde{N}_w}\sum_{i:\tilde{W}_i=w}\tilde{Y}_i$ and $B_w \sim \mathrm{Binomial}(N-1, \rho_w)$ for $w = 0, 1$.

In practice, the expectation $\mathbb{E}\left[1/(B+1)^2\right]$ can be easily approximated by the Monte Carlo simulation.

Using the variance estimator and assuming the asymptotic normality of the DM estimator, we can construct the nominal central confidence interval at the significance level α as:

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

G.2 OLS estimator

The DM/IPW estimators are widely accepted frequentist estimators for their simplicity. Another important estimator is the OLS estimator. This section discusses the OLS estimator for τ in a scenario where all variables, including covariates, are jointly privatized. The observed covariates are privatized by the Laplace mechanism. We assume $X_{i,j} \in [0,1]$ for $i = 1, \dots, N$ and $j = 1, \dots, d$ to ensure bounded ℓ_1-sensitivity. The privatized outcomes and covariates are

$$\tilde{X}_{i,j} = X_{i,j} + \nu^X_{i,j},$$

where

$$\nu^X_{i,j} \overset{i.i.d}{\sim} \mathrm{Lap}(d/\epsilon_x).$$

By composition, the joint release of $(\tilde{Y}_i, \tilde{X}_{i,1}, \dots, \tilde{X}_{i,d}, \tilde{W}_i)_{i=1}^N$ 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 (Lei & Ding, 2020). Specifically, we propose the following plug-in OLS estimator.

$$\tilde{\tau}_{OLS} = \tilde{\alpha}_{(1)} - \tilde{\alpha}_{(0)} + \bar{X}(\tilde{\beta}_{(1)} - \tilde{\beta}_{(0)}), \tag{10}$$

where $\bar{X} = \frac{1}{N}\sum_{i=1}^N \tilde{X}_i$ and

$$(\tilde{\alpha}_{(w)}, \tilde{\beta}_{(w)}) = \underset{\alpha,\beta}{\arg\min} \sum_{i:\tilde{W}_i=w}(\tilde{Y}_i - \alpha - \tilde{X}_i'\beta)^2$$

for $w = 0, 1$. Note that, under some regularity conditions (Lehmann & Casella, 1998, p. 440), $(\tilde{\alpha}_{(w)}, \tilde{\beta}_{(w)})$ converges to $(\tilde{\alpha}^*_{(w)}, \tilde{\beta}^*_{(w)})$, defined as

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

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 (10) 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 G.5. *The estimator $C_{p,\epsilon_w}\tilde{\tau}_{OLS}$ is consistent for τ. Furthermore, $\sqrt{N}(C_{p,\epsilon_w}\tilde{\tau}_{OLS} - \tau)$ converges in distribution to*

$$\mathrm{N}\left(0, C^2_{p,\epsilon_w}\left(\frac{MSE_1}{\rho_1} + \frac{MSE_0}{\rho_0}\right)\right), \tag{11}$$

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

We adopt the plug-in estimator for the asymptotic variance in (11), that is, we let:

$$\widehat{MSE}_w = \frac{1}{\tilde{N}_w} \sum_{i:\tilde{W}_i = w} (\tilde{Y}_i - \tilde{\alpha}_{(w)} - \tilde{X}_i \tilde{\beta}_{(w)})^2$$

for $w = 0, 1$, which is a consistent estimator for MSE_w. Our asymptotically valid confidence interval at significance level α is:

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

where $\hat{\Sigma}_{OLS} = C_{p,\epsilon_w}^2 \left(\frac{\widehat{MSE}_1}{\rho_1} + \frac{\widehat{MSE}_0}{\rho_0} \right)$.

H Proofs

H.1 Proof of Lemma 4.1

Proof. We first acknowledge that

$$\sup_{\substack{P_0 = \delta(0), \\ P_1 \in \mathcal{P}_1, \\ p=1}} \mathbb{E}[(\hat{\tau} - \tau)^2] = \sup_{P_1 \in \mathcal{P}_1} \mathbb{E}[(\hat{\tau} - \mu_1)^2], \tag{12}$$

where $\delta(0)$ denotes a point mass at 0 and $\mu_1 = \mathbb{E}[Y_i(1)]$. Equation (12) 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. □

H.2 Proof of Lemma G.1

Proof. We let $\bar{p} = 1 - p$ and $\bar{q}_{\epsilon_w} = 1 - q_{\epsilon_w}$ throughout the proofs. Let $\tilde{A}_k = \{\tilde{W}_1 = ... = \tilde{W}_k = 1, \tilde{W}_{k+1} = ... = \tilde{W}_N = 0\}$ for $k = 1, .., N-1$. The event $\tilde{A}_k$ is permutation invariant for all $\tilde{W}_i$. That is, all the events that have k ones and $N - k$ zeros are equally likely. Also, $\tilde{A}_j$ and $\tilde{A}_k$ are mutually exclusive for $j \neq k$.

$$\mathbb{E}[\tilde{\tau}_{DM}] = \sum_{k=1}^{N-1} \binom{N}{k} \mathbb{E}[\tilde{\tau}_{DM} \mid \tilde{A}_k] p(\tilde{A}_k), \tag{13}$$

where

$$
\begin{aligned}
\mathbb{E}[\tilde{\tau}_{DM} \mid \tilde{A}_k] =& \mathbb{E}\Bigg[\frac{1}{k}\sum_{i=1}^{N}\tilde{W}_i\tilde{Y}_i - \frac{1}{N-k}\sum_{i=1}^{N}(1-\tilde{W}_i)\tilde{Y}_i \mid A_k\Bigg] \\
=& \mathbb{E}[\tilde{Y} \mid \tilde{W}_i = 1] - \mathbb{E}[Y_i \mid \tilde{W}_i = 0] \\
=& \sum_{w=0}^{1}\mathbb{E}[\tilde{Y}_i \mid W_i = w, \tilde{W}_i = 1]p(W_i = w \mid \tilde{W}_i = 1) - \sum_{w=0}^{1}\mathbb{E}[\tilde{Y}_i \mid W_i = w, \tilde{W}_i = 0]p(W_i = w \mid \tilde{W}_i = 0) \\
=& \sum_{w=0}^{1}\mathbb{E}[Y_i \mid W_i = w, \tilde{W}_i = 1]p(W_i = w \mid \tilde{W}_i = 1) - \sum_{w=0}^{1}\mathbb{E}[Y_i \mid W_i = w, \tilde{W}_i = 0]p(W_i = w \mid \tilde{W}_i = 0) \\
=& \frac{\bar{p}\bar{q}_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}}\mathbb{E}[Y_i(0)] + \frac{pq_{\epsilon_w}}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}}\mathbb{E}[Y_i(1)] - \frac{\bar{p}q_{\epsilon_w}}{\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w}}\mathbb{E}[Y_i(0)] - \frac{p\bar{q}_{\epsilon_w}}{\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w}}\mathbb{E}[Y_i(1)] \\
=& \frac{(q_{\epsilon_w} - \bar{q}_{\epsilon_w})p\bar{p}}{(\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w})(pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w})}(\mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]) \\
=& \frac{(q_{\epsilon_w} - \bar{q}_{\epsilon_w})p\bar{p}}{(\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w})(pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w})}\tau,
\end{aligned}
$$

for $k = 1, ..., N-1$. The third line follows from the law of total expectation, the fourth line follows from the independence of DP noise and the fifth line follows from Lemma B.1 and SUTVA.

Also notice that

$$
p(\{\tilde{W}_1 = ... = \tilde{W}_N = 0\}) = (\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w})^N = \rho_0^N,
$$

and

$$
p(\{\tilde{W}_1 = ... = \tilde{W}_N = 1\}) = (pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w})^N = \rho_1^N.
$$

Putting altogether, we prove our claim. □

H.3 Proof of Lemma G.2

We prove the consistency of the estimator.

Proof. Note that

$$
\tilde{\tau}_{DM} = \frac{N}{\tilde{N}_1}\frac{1}{N}\sum_{i=1}^{N}\tilde{W}_i\tilde{Y}_i - \frac{N}{\tilde{N}_0}\frac{1}{N}\sum_{i=0}^{N}(1-\tilde{W}_i)\tilde{Y}_i.
$$

Also,

$$
\frac{N}{\tilde{N}_1} \xrightarrow{p} \frac{1}{p(\tilde{W}_i = 1)} = \frac{1}{pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}},
$$

and 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 B.1. Similarly, we have

$$\frac{N}{\tilde{N}_0} \xrightarrow{p} \frac{1}{p(\tilde{W}_i = 0)} = \frac{1}{\bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w}},$$

and

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

Therefore, by the continuous mapping theorem, we can see that

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

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

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

□

H.4 Proof of Lemma G.3

Proof. First, note that

$$\mathbb{V}\mathrm{ar}[C_{p,\epsilon_w}\tilde{\tau}_{DM}] = C_{p,\epsilon_w}^2\mathbb{V}\mathrm{ar}[\tilde{\tau}_{DM}].$$

By the law of total variance, we have

$$\mathbb{V}\mathrm{ar}[\tilde{\tau}_{DM}] = \mathbb{V}\mathrm{ar}[\mathbb{E}[\tilde{\tau}_{DM} \mid \tilde{\mathbf{W}}]] + \mathbb{E}[\mathbb{V}\mathrm{ar}[\tilde{\tau}_{DM} \mid \tilde{\mathbf{W}}]], \tag{14}$$

where $\tilde{\mathbf{W}} = \{\tilde{W}_1, ..., \tilde{W}_N\}$. Note that,

$$\mathbb{E}[\tilde{\tau}_{DM} \mid \tilde{\mathbf{W}}] = \begin{cases} 0 \text{ if } \tilde{\mathbf{W}} = \mathbf{0}_N \text{ or } \mathbf{1}_N \\ \frac{1}{C_{p,\epsilon_w}}\tau \text{ otherwise,} \end{cases}$$

where $\mathbf{0}_N$ and $\mathbf{1}_N$ denote N-dimensional the zero and one vectors. Let $U = \mathbb{E}[\tilde{\tau}_{DM} \mid \tilde{\mathbf{W}}]$. Then, the first term of (14) can be written as

$$\mathbb{V}\mathrm{ar}[U] = \mathbb{E}[U^2] - \mathbb{E}[U]^2 = (1 - r^N - (1-r)^N)(r^N + (1-r)^N)\frac{\tau^2}{C_{p,\epsilon_w}^2},$$

where $r = pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}$.

The second term of (14) can be equivalently written as

$$\mathbb{E}[\mathbb{V}\mathrm{ar}[\tilde{\tau}_{DM} \mid \tilde{\mathbf{W}}]] = \sum_{k=1}^{N-1}\binom{N}{k}\mathbb{V}\mathrm{ar}[\tilde{\tau}_{DM} \mid \tilde{A}_k]p(\tilde{A}_k). \tag{15}$$

Note that

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\tilde{\tau}_{DM} \mid \tilde{A}_k] &= \mathbb{V}\mathrm{ar}\left[\frac{1}{\tilde{N}_1}\sum_{i=1}^{N}\tilde{W}_i\tilde{Y}_i - \frac{1}{\tilde{N}_0}\sum_{i=1}^{N}(1-\tilde{W}_i)\tilde{Y}_i \mid \tilde{A}_k\right] \\
&= \mathbb{V}\mathrm{ar}\left[\frac{1}{k}\sum_{i=1}^{k}\tilde{W}_i\tilde{Y}_i - \frac{1}{N-k}\sum_{i=k+1}^{N}(1-\tilde{W}_i)\tilde{Y}_i \mid \tilde{A}_k\right] \\
&= \frac{1}{k}\mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] + \frac{1}{N-k}\mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 0] \\
&= \frac{1}{k}\mathbb{V}\mathrm{ar}[Y_i \mid \tilde{W}_i = 1] + \frac{1}{N-k}\mathbb{V}\mathrm{ar}[Y_i \mid \tilde{W}_i = 0] + \frac{2N}{k(N-k)\epsilon_y^2}.
\end{aligned}$$

The third line follows from the i.i.d assumption. By the law of total variance and SUTVA,

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

The first term simplifies to

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

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

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

Letting $\rho_1 = pq_{\epsilon_w} + \bar{p}\bar{q}_{\epsilon_w}$ and $\rho_0 = \bar{p}q_{\epsilon_w} + p\bar{q}_{\epsilon_w}$, we obtain the exact variance.

Finally, we consider the asymptotic order of the variance. It is straightforward that $(1 - \rho_1^N - \rho_0^N)(\rho_1^N + \rho_0^N)\tau^2$ has the order of $O((pq_{\epsilon_w})^N)$, therefore we only consider the order of $C_{p,\epsilon_w}^2 \sum_{k=1}^{N-1} \binom{N}{k} \rho_1^k \rho_0^{N-k} \left\{ \frac{V_1}{k} + \frac{V_0}{N-k} + \frac{2N}{k(N-k)\epsilon_y^2} \right\}$. First, we have

$$C_{p,\epsilon_w}^2 = O((\epsilon_w^2)^{-1}).$$

Also, Lemma C.1 implies that

$$\sum_{k=1}^{N-1} \binom{N}{k} \rho_1^k \rho_0^{N-k} \frac{1}{k} = O\left(\frac{1}{Npq_{\epsilon_w}} \right).$$

Thus, we have

$$C_{p,\epsilon_w}^2 \sum_{k=1}^{N-1} \binom{N}{k} \rho_1^k \rho_0^{N-k} \frac{V_1}{k} = O\left(\frac{q_{\epsilon_w}}{Np} \right) = O\left(N^{-1}\right), \quad C_{p,\epsilon_w}^2 \sum_{k=1}^{N-1} \binom{N}{k} \rho_1^k \rho_0^{N-k} \frac{V_0}{N-k} = O\left(\frac{q_{\epsilon_w}}{Np} \right) = O\left(N^{-1}\right).$$

Also note that,

$$C_{p,\epsilon_w}^2 \sum_{k=1}^{N-1} \binom{N}{k} \rho_1^k \rho_0^{N-k} \frac{2N}{k(N-k)\epsilon_y^2} \leq C_{p,\epsilon_w}^2 \sum_{k=1}^{N-1} \binom{N}{k} \rho_1^k \rho_0^{N-k} \frac{2N}{k^2\epsilon_y^2} = O\left(\frac{1}{N\epsilon_y^2\epsilon_w^2} \right).$$

Therefore, we have

$$C_{p,\epsilon_w}^2 \sum_{k=1}^{N-1} \binom{N}{k} \rho_1^k \rho_0^{N-k} \left\{ \frac{V_1}{k} + \frac{V_0}{N-k} + \frac{2N}{k(N-k)\epsilon_y^2} \right\} = O\left((N\epsilon_y^2\epsilon_w^2)^{-1}\right),$$

which proves our claim. □

H.5 Proof of Lemma G.4

Proof.

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

Therefore,

$$\begin{aligned}
\mathbb{E}[\hat{s}_1^2 \mid \tilde{W}_i = 1] &= \frac{\tilde{N}_1}{\tilde{N}_1 - 1} \mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] - \frac{\tilde{N}_1}{\tilde{N}_1 - 1} \mathbb{V}\mathrm{ar}[\bar{Y}_i \mid \tilde{W}_i = 1] \\
&= \frac{\tilde{N}_1}{\tilde{N}_1 - 1} \mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] - \frac{1}{\tilde{N}_1 - 1} \mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] \\
&= \mathbb{V}\mathrm{ar}[\tilde{Y}_i \mid \tilde{W}_i = 1] \\
&= \mathbb{V}\mathrm{ar}[Y_i \mid \tilde{W}_i = 1] + \frac{2}{\epsilon_y^2}.
\end{aligned}$$

Therefore,

$$\mathbb{E}[\hat{s}_1^2 - \frac{2}{\epsilon_y^2} \mid \tilde{W}_i = 1] = \mathbb{V}\mathrm{ar}[Y_i \mid \tilde{W}_i = 1] = V_1.$$

We can follow the same procedure for $\mathbb{E}[\hat{s}_0^2 - \frac{2}{\epsilon_y^2}] = V_0$. By plugging these results into Lemma G.3, we can find that the following estimator $\hat{\mathbb{V}}_{DM}$ is an asymptotically unbiased estimator for $\mathbb{V}\mathrm{ar}[C_{p,\epsilon_w} \tilde{\tau}_{DM}]$:

$$\hat{\mathbb{V}}_{DM} = C_{p,\epsilon_w}^2 \sum_{k=1}^{N-1} \binom{N}{k} \rho_1^k \rho_0^{N-k} \left\{ \frac{1}{k} \hat{s}_1^2 + \frac{1}{N-k} \hat{s}_0^2 \right\}. \tag{16}$$

Finally, we apply a Binomial approximation technique that is provided in Section I. □

H.6 Proof of Lemma 4.2

See the proof in Section H.3.

H.7 Proof of Theorem D.1

Proof. Consistency is proven in a similar fashion as in Section H.3.

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

Note that $\tilde{\tau}_i$ is i.i.d. for $i = 1, \dots, N$, $\mathbb{E}[C_{p,\epsilon_w} \tilde{\tau}_i] = \tau$, 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 $\mathbb{V}\mathrm{ar}[C_{p,\epsilon_w} \tilde{\tau}_i] = C_{p,\epsilon_w}^2 \mathbb{V}\mathrm{ar}[\tilde{\tau}_i]$.

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

Then,

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

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

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

Putting all together, we prove the central limit theorem in Theorem D.1.

Finally, note that Section H.4 provides the decomposition of V_w, and

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

which follows from Lemma B.1.

□

H.8 Proof of Lemma D.2

See the proof in Section H.4.

H.9 Proof of Corollary D.3

See the proof in Section H.4.

H.10 Proof of Theorem F.1

First, we have

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

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 F.1. The estimator of Σ^* that we adopt in Section 4.4 are a plug-in estimator with consistent estimators of $\nabla h(\mathbf{E})$ and S^*.

H.11 Proof of Theorem G.5

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

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

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

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

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

$$\begin{aligned} MSE_w &= \mathbb{E}[(\tilde{Y}_i - \tilde{\gamma}^*_w - (\tilde{X}_i - \mu_{\tilde{X}})\tilde{\beta}^*_w)^2 \mid \tilde{W}_i = w] \\ &= \mathbb{E}[(\tilde{Y}_i - \tilde{\alpha}^*_w - \tilde{X}_i'\tilde{\beta}^*_w)^2 \mid \tilde{W}_i = w]. \end{aligned}$$

Therefore, we have

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

which implies

$$\sqrt{N}(\tilde{\gamma}_{(w)} - \tilde{\gamma}^*_{(w)}) \xrightarrow{D} \mathrm{N}\left(0, \frac{MSE_w}{p(\tilde{W}_i = w)}\right). \tag{17}$$

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

$$\sqrt{N}(C_{p,\epsilon_w}\hat{\tau}_{OLS} - \tau) \xrightarrow{D} \mathrm{N}\left(0, C^2_{p,\epsilon_w}\left(\frac{MSE_1}{\rho_1} + \frac{MSE_0}{\rho_0}\right)\right).$$

□

I Approximation of the Expectation of Inverse Binomial

When evaluating Equation 16, we need to evaluate

$$\sum_{k=1}^{N-1} \frac{1}{k}\binom{N}{k} p^k (1-p)^{N-k},$$

which is computationally infeasible when N is large. We can express the summation in the following way.

$$\begin{aligned} &\sum_{k=1}^{N-1} \frac{1}{k}\binom{N}{k} p^k (1-p)^{N-k} \\ &= \sum_{k=1}^{N-1} \frac{N}{k^2}\binom{N-1}{k-1} p^k (1-p)^{N-k} \\ &= Np \sum_{k=1}^{N-1} \frac{1}{k^2}\binom{N-1}{k-1} p^{k-1} (1-p)^{N-k} \\ &= Np \left\{ \sum_{k=1}^{N} \frac{1}{k^2}\binom{N-1}{k-1} p^{k-1} (1-p)^{N-k} - \frac{p^{N-1}}{N^2} \right\} \\ &= Np \left\{ \sum_{k=0}^{N-1} \frac{1}{(k+1)^2}\binom{N-1}{k} p^{k} (1-p)^{N-k-1} - \frac{p^{N-1}}{N^2} \right\} \\ &= Np\mathbb{E}\left[\frac{1}{(B+1)^2}\right] - \frac{p^N}{N}, \end{aligned}$$

where $B \sim \text{Binomial}(N-1, p)$. The moment $\mathbb{E}[\frac{1}{(B+1)^2}]$ can be easily approximated by the Monte Carlo simulation.

J Bayesian Methodology

J.1 Details of the DPM

We say the probability measure H is generated from a Dirichlet Process, DP(αH_0), with a concentration parameter $\alpha > 0$ and a base probability measure H_0 over a measurable space $(\Theta, \mathcal{B})$ (Ferguson, 1974) if, for any finite partition $(B_1, ..., B_k)$ of $\mathcal{B}$, we have

$$\big(H(B_1), ..., H(B_k)\big) \sim \text{Dir}\big(\alpha H_0(B_1), ..., \alpha H_0(B_k)\big),$$

where 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 $\mathbf{\Phi} = (\Phi_1, ..., \Phi_N)$ can be rewritten in terms of its unique values $\mathbf{\Phi}^* = (\Phi_1^*, ..., \Phi_K^*)$. In particular, this can be represented in the following stick-breaking process.

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

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

$$p(Y_i(w)|\boldsymbol{\mu}, \mathbf{\Sigma}) \propto \sum_{k=1}^{\infty} u_k \text{N}(\mu_w^k, \Sigma_w^k), \tag{18}$$

where 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 DP(αH_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 & 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 are potentially belonged to by units. 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 $C_i = k$, (18) is greatly simplified to

$$p(Y_i(w)|\boldsymbol{\mu}, \mathbf{\Sigma}) \propto \text{N}(\mu_w^k, \Sigma_w^k).$$

Ishwaran & 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.

J.2 Detailed Steps of Gibbs Sampler

In this section we present the detailed steps of the Gibbs sampler that is described in Section 5.1. The algorithm is inspired by Schwartz et al. (2011) and Ohnishi & Sabbaghi (2022b). First, we present the outline of the Gibbs sampler.

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:
$$p(Y_i(W_i)|-) \propto p(Y_i(W_i) \mid \mu_{W_i}^{C_i}, \Sigma_{W_i}^{C_i})p(\tilde{Y}_i \mid Y_i(W_i))$$
$$p(Y_i(1-W_i)|-) \propto p(Y_i(1-W_i) \mid \mu_{1-W_i}^{C_i}, \Sigma_{1-W_i}^{C_i}).$$
The Privacy-Aware Metropolis-within-Gibbs algorithm Ju et al. (2022) is used for this step.
3. Given $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{u}$, $Y_i(0)$ and $Y_i(1)$, draw each C_i from
$$p(C_i = k|-) \propto u_k p(Y_i(0) \mid \mu_0^k, \Sigma_0^k)p(Y_i(1) \mid \mu_1^k, \Sigma_1^k).$$
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 normal 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
$$p(\mu_0^k, \Sigma_0^k|-) \propto H(\mu_0^k, \mu_1^k, \Sigma_0^k, \Sigma_1^k) \prod_{i:C_i=k} p(Y_i(0), Y_i(1) \mid \mu_0^k, \mu_1^k, \Sigma_0^k, \Sigma_1^k).$$

The details of each step are 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, for unit i with $\tilde{W}_i = 0$,
$$r_0 = \mathrm{Lap}(\tilde{\mathrm{Y}}_{\mathrm{i}} \mid \mathrm{Y}_{\mathrm{i}}(0), 1/\epsilon_{\mathrm{y}})q_{\epsilon_w}(1-p) \text{ and } r_1 = \mathrm{Lap}(\tilde{\mathrm{Y}}_{\mathrm{i}} \mid \mathrm{Y}_{\mathrm{i}}(1), 1/\epsilon_{\mathrm{y}})(1-q_{\epsilon_w})p,$$
and for unit i with $\tilde{W}_i = 1$,
$$r_0 = \mathrm{Lap}(\tilde{\mathrm{Y}}_{\mathrm{i}} \mid \mathrm{Y}_{\mathrm{i}}(0), 1/\epsilon_{\mathrm{y}})(1-q_{\epsilon_w})(1-p) \text{ and } r_1 = \mathrm{Lap}(\tilde{\mathrm{Y}}_{\mathrm{i}} \mid \mathrm{Y}_{\mathrm{i}}(1), 1/\epsilon_{\mathrm{y}})q_{\epsilon_w}p.$$
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 $\alpha = \min\left(1, \frac{\mathrm{Lap}(\mathrm{y}*|\tilde{\mathrm{Y}}_{\mathrm{i}}, 1/\epsilon_{\mathrm{y}})}{\mathrm{Lap}(\mathrm{y}^{\mathrm{prev}}|\tilde{\mathrm{Y}}_{\mathrm{i}}, 1/\epsilon_{\mathrm{y}})}\right)$,

   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
$$p(u'_k|-) \propto \text{Beta}\bigg(1 + \sum_{i:C_i=k} 1, \alpha + \sum_{i:C_i>k} 1\bigg).$$
Then, update $u_k = u'_k \prod_{j<k}(1 - u'_j)$.
5. Given $\mathbf{C}$ and $\mathbf{u}'$, draw α from
$$p(\alpha|-) \propto p(\alpha) \prod_{k=1}^{K} f\bigg(u'_k\bigg|1 + \sum_{i:C_i=k} 1, \alpha + \sum_{i:C_i>k} 1\bigg),$$
where f is the pdf of u'_k, the beta distribution. The Metropolis-Hastings algorithm is used for this step with a proposal distribution $\text{TN}(\alpha^{prev}, 1.0, 0, \infty)$. 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^k_w from $\text{IG}(2 + 0.5N_k, 0.2^2 + 0.5s^k_w)$ where $s^k_w = \sum_{i:C_i=k}(Y_i(w) - \mu^k_w)^2$ for $w = 0, 1$. If $N_k = 0$,then draw Σ^k_w from the prior $\text{IG}(2, 0.2^2)$.

   (b) If $N_k > 0$, draw μ^k_w from
   $$\text{TN}\bigg(\frac{0.5 * \Sigma^k_w + 9.0s_w}{\Sigma^k_w + 9.0N_k}, \frac{9.0\Sigma^k_w}{\Sigma^k_w + 9.0N_k}, 0, 1\bigg),$$
   where $s_w = \sum_{i=1}^{N} Y_i(w)$. If $N_k = 0$, draw μ^k_w from
   $$\text{TN}(0.5, 9.0, 0, 1).$$
   We use a common choice of the base measure H_0: the Normal-Inverse-Gamma conjugate $\text{N}(\mu_0, \sigma_0^2)\text{N}(\mu_0, \sigma_0^2)\text{IG}(a_0, b_0)\text{IG}(a_0, b_0)$. 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 $w = 0, 1$.

J.3 Modifications for Custom Scenario in Section 4.3

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
$$r_w = p(\tilde{A}_i \mid Y_i(w), W_i = w)p(W_i = w),$$
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:
$$p(Y_i(W_i)|-) \propto p(Y_i(W_i) \mid \mu^{C_i}_{W_i}, \Sigma^{C_i}_{W_i})p(\tilde{A}_i \mid Y_i(W_i))$$
$$p(Y_i(1 - W_i)|-) \propto p(Y_i(1 - W_i) \mid \mu^{C_i}_{1-W_i}, \Sigma^{C_i}_{1-W_i}).$$
The Privacy-Aware Metropolis-within-Gibbs algorithm Ju et al. (2022) is used for this step.

J.4 Modifications for Custom Scenario in Section 4.4

Under the custom scenario in Section 4.4, we do not have access to p. Therefore, we need an additional step to infer p. Specifically, with a prior distribution $p \sim \text{Beta}(1, 1)$, we add the following step.

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

The other steps are similar to J.3.

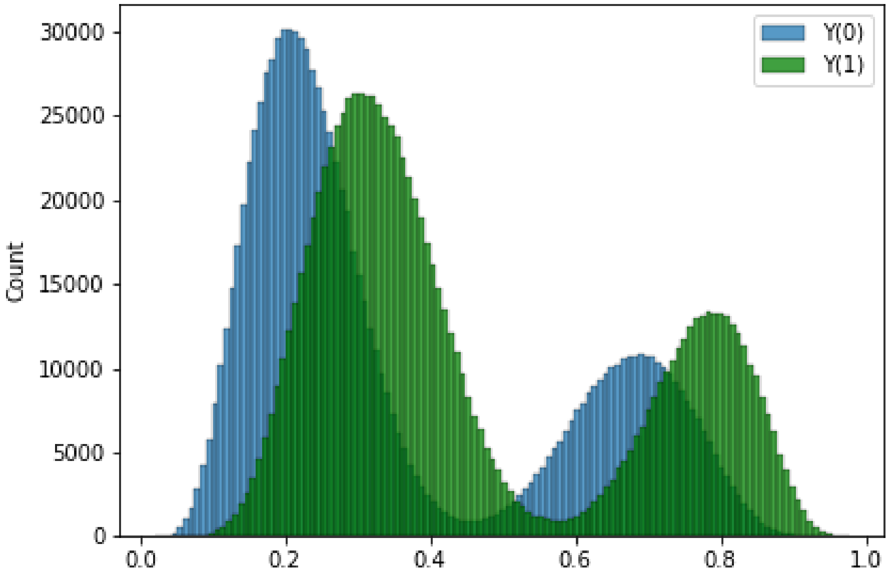

Figure 1: Distributions of $Y(0)$ and $Y(1)$ for simulation studies.

K Beta GLM

Under the data-generating processes and the re-parameterizations of the Beta regression provided in Section 6.1, we generated 1000000 samples for $Y(0)$ and $Y(1)$ to see what the data looks like. Figure 1 shows the distributions of each potential outcome. Also, the expectations of each potential outcome are expressed as:

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

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

L Additional simulations

L.1 Additional simulation results for Bayes estimators

We further investigated the performance of our Bayesian approach under different data-generating mechanisms. The data is generated from the same data-generating mechanisms as in 6.1 with different parameterizations, $\mu_i(w) = \text{expit}(1.0 - 1.3X_1 + 1.5X_2 - 2.0X_3 + 1.0w)$. We observe that the Bayesian estimators exhibit good bias and MSE at the cost of less-calibrated probabilities for $\epsilon_{\text{tot}} = 0.3, 3$. Figure 2 shows the distributions of each potential outcome.

L.2 Simulation results for the DM estimator

Table 4 and 5 present the simulation results for the DM estimator. The data are generated according to the mechanisms provided in Section 6.1. Note that these results are based on the assumption of the asymptotic normality of the estimator.

Table 3: Evaluation metrics for Bayes estimators ($N = 1000$, $N_{sim} = 200$).

		Coverage		Bias		MSE		Interval Width	
ϵ_{tot}	$(\epsilon_x, \epsilon_y, \epsilon_w)$	OLS	Bayes	OLS	Bayes	OLS	Bayes	OLS	Bayes
0.3	$(0.1, 0.1, 0.1)$	93.0%	76.0%	-0.390	-0.210	1.100	0.0443	1.860	0.469
3	$(1, 1, 1)$	95.0%	76.0%	0.0132	-0.123	0.0406	0.0188	0.763	0.347
9	$(3, 3, 3)$	97.0%	93.0%	-0.00323	-0.0153	0.000949	0.000938	0.135	0.107
30	$(10, 10, 10)$	96.0%	93.0%	0.000845	-0.00464	0.000172	0.000183	0.0504	0.0482

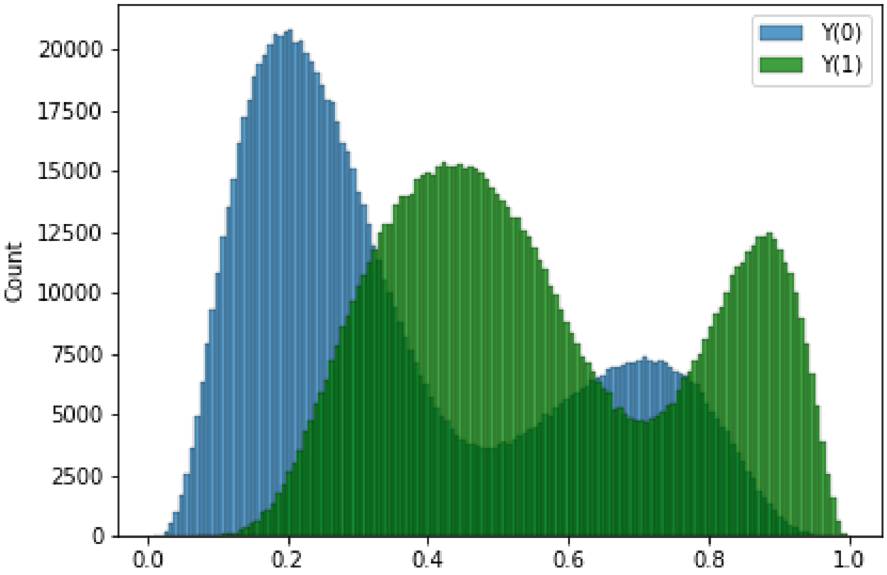

Figure 2: Distributions of $Y(0)$ and $Y(1)$ for an additional simulation study.

Table 4: Evaluation metrics for DM and IPW estimators ($N = 1000$, $N_{sim} = 2000$). 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}	$(\epsilon_x, \epsilon_y, \epsilon_w)$	DM	IPW	DM	IPW	DM	IPW	DM	IPW
3	$(1, 1, 1)$	95.2%	95.3%	-0.00344	-0.00266	0.0369	0.0405	0.767	0.798
9	$(3, 3, 3)$	96.1%	95.4%	-0.000442	-0.00105	0.00127	0.00208	0.142	0.181
30	$(10, 10, 10)$	95.9%	95.0%	-0.000656	-0.000547	0.000265	0.000906	0.065	0.120
0.3	$(0.1, 0.1, 0.1)$	95.3%	95.5%	-0.130	-0.129	0.986	0.989	1.905	1.909
3	$(2, 0.5, 0.5)$	94.4%	94.5%	-0.00890	-0.00703	0.375	0.378	1.746	1.748
3	$(0.5, 2, 0.5)$	95.2%	95.2%	-0.00414	-0.00576	0.0372	0.0484	0.751	0.857
3	$(0.5, 0.5, 2)$	94.9%	95.4%	-0.00240	-0.00238	0.0567	0.0575	0.920	0.929
3	$(0.5, 1.25, 1.25)$	94.5%	94.6%	0.00277	0.00480	0.0186	0.0210	0.515	0.547
3	$(1.25, 0.5, 1.25)$	95.3%	95.3%	-0.00178	-0.00101	0.101	0.103	1.222	1.232
3	$(1.25, 1.25, 0.5)$	95.2%	94.6%	0.00175	0.00137	0.0887	0.102	1.141	1.195

Table 5: Evaluation metrics for DM and IPW estimators ($N = 10000$, $N_{sim} = 2000$).

		Coverage		Bias		MSE		Interval Width	
ϵ_{tot}	$(\epsilon_x, \epsilon_y, \epsilon_w)$	DM	IPW	DM	IPW	DM	IPW	DM	IPW
3	$(1, 1, 1)$	94.6%	95.4%	-0.000587	-0.00174	0.00401	0.00407	0.243	0.252
9	$(3, 3, 3)$	95.7%	94.7%	-0.000055	-0.000154	0.000128	0.000216	0.0451	0.0573
30	$(10, 10, 10)$	95.9%	94.6%	-0.000156	0.000213	0.0000261	0.0000962	0.0205	0.0380
0.3	$(0.1, 0.1, 0.1)$	95.2%	94.4%	-0.125	-0.104	0.922	0.919	1.899	1.883
3	$(2, 0.5, 0.5)$	94.4%	94.9%	-0.00516	-0.00380	0.0529	0.0535	0.905	0.915
3	$(0.5, 2, 0.5)$	94.6%	95.7%	0.00124	0.00112	0.00378	0.00466	0.237	0.271
3	$(0.5, 0.5, 2)$	95.8%	95.9%	0.000399	0.000703	0.00566	0.00524	0.292	0.295
3	$(0.5, 1.25, 1.25)$	95.3%	95.9%	-0.00126	0.00133	0.00173	0.00187	0.163	0.173
3	$(1.25, 0.5, 1.25)$	94.9%	95.1%	-0.000977	-0.000968	0.0105	0.0106	0.401	0.405
3	$(1.25, 1.25, 0.5)$	95.5%	95.4%	-0.000990	0.00247	0.00860	0.00957	0.369	0.391

Table 6: Evaluation metrics for IPW and OLS estimators ($N = 1000, N_{sim} = 2000$). 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}	$(\epsilon_x, \epsilon_y, \epsilon_w)$	IPW	OLS	IPW	OLS	IPW	OLS	IPW	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

L.3 Evaluation of OLS estimator

We evaluate the frequentist properties of the OLS estimators. We consider the standard privacy mechanism in Section 4.2 and use the same data-generating mechanisms in Section 6. 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 $\text{Lap}(3/\epsilon_y)$ 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.

L.3.1 Results

Tables 6 and 7 present the performance evaluation of the IPW 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 8 compares our Bayesian methodology with the OLS estimator for $N = 1000$. We see that the Bayes estimator yields well-calibrated coverage probabilities, and achieves the same level of MSE for conservative budgets, e.g., $(\epsilon_x, \epsilon_y, \epsilon_w) = (3, 3, 3), (10, 10, 10)$. The bias is larger than that of the OLS estimator because the Bayes estimator is biased in finite samples because of the priors. Under strict budgets, however, the Bayes estimator outperforms the OLS estimator in terms of MSE and bias at the cost of over-coverage. More importantly, the interval length of the Bayes estimator for $(\epsilon_x, \epsilon_y, \epsilon_w) = (0.1, 0.1, 0.1)$ is 0.468, which is informative. One limitation of the Bayesian approach is that the coverage of the posterior interval is not necessarily well-calibrated. We observe a less-calibrated coverage probability, 100%, for $(\epsilon_x, \epsilon_y, \epsilon_w) = (0.1, 0.1, 0.1)$. We also provide simulation results under the same data-generating mechanisms as in 6.1 with different parameters, in which the Bayesian posterior intervals exhibit under-coverage rather than over-coverage.

L.3.2 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 IPW 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

Table 7: Evaluation metrics for IPW and OLS estimators ($N = 10000, N_{sim} = 2000$).

		Coverage		Bias		MSE		Interval Width	
ϵ_{tot}	$(\epsilon_x, \epsilon_y, \epsilon_w)$	IPW	OLS	IPW	OLS	IPW	OLS	IPW	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

Table 8: Evaluation metrics for Bayes estimators ($N = 1000, N_{sim} = 200$).

		Coverage		Bias		MSE		Interval Width	
ϵ_{tot}	$(\epsilon_x, \epsilon_y, \epsilon_w)$	OLS	Bayes	OLS	Bayes	OLS	Bayes	OLS	Bayes
0.3	$(0.1, 0.1, 0.1)$	96.0%	100.0%	-0.142	-0.0976	0.981	0.00973	1.911	0.468
3	$(1, 1, 1)$	95.5%	96.0%	-0.0404	-0.0707	0.0340	0.00863	0.769	0.374
9	$(3, 3, 3)$	95.5%	93.0%	0.00107	-0.0121	0.00139	0.00105	0.141	0.115
30	$(10, 10, 10)$	98.0%	97.0%	0.000339	-0.00445	0.000152	0.000167	0.0581	0.0552

IPW 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 IPW estimator in terms of MSE. This result follows from the fact that the regression adjustment technique in randomized experiments (Freedman, 2008; Lei & Ding, 2020) 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 briefly discuss some limits to the gains in precision of the estimator for the PATE from including covariates in our scenario. In large samples, including covariates in the regression function under usual randomized experiments will not lower the precision (Lei & Ding, 2020; Imbens & Rubin, 2015). However, DP mechanisms under randomization pose unique challenges. We first note that MSE_w in Theorem G.5 can be written as follows:

$$\begin{aligned} MSE_w = {} & \mathbb{V}\text{ar}[Y_i|\tilde{W}_i = w] + \mathbb{E}[Y_i|\tilde{W}_i = w]^2 + \frac{1}{\epsilon_y^2} \\ & - \mathbb{E}[\tilde{Y}_i'\tilde{X}_i(\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i'\tilde{Y}_i|\tilde{W}_i = w]. \end{aligned} \tag{19}$$

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 (19) 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. In support of this observation, we see that the gain in MSE for the allocation $(\epsilon_x, \epsilon_y, \epsilon_w) = (1.25, 1.25, 0.5)$ is comparatively substantial among all other budget allocations with $\epsilon_{tot} = 3$ because ϵ_x and ϵ_y are both small. Future researchers may investigate precise conditions for when the variance reduction of the OLS estimator out-weighs the additional privacy cost of $\tilde{X}$, as well as general principles for privacy budget allocation.