Optimal Federated Learning for Functional Mean Estimation under Heterogeneous Privacy Constraints

T. Tony Cai, Abhinav Chakraborty, Lasse Vuursteen

June 2, 2026

Abstract

Federated learning (FL) is a distributed machine learning technique designed to preserve data privacy and security, and it has gained significant importance due to its broad range of applications. This paper addresses the problem of optimal functional mean estimation from discretely sampled data in a federated setting.

We consider a heterogeneous framework where the number of individuals, measurements per individual, and privacy parameters vary across one or more servers, under both common and independent design settings. In the common design setting, the same design points are measured for each individual, whereas in the independent design setting, each individual has their own random collection of design points. Within this framework, we establish minimax upper and lower bounds for the estimation error of the underlying mean function, highlighting the differences between common and independent designs under distributed privacy constraints.

We propose algorithms that achieve the optimal trade-off between privacy and accuracy and provide optimality results that quantify the fundamental limits of private functional mean estimation. We further support the theory with simulations and a real-data illustration using BMI trajectories from the Health and Retirement Study.

Keywords: Differential Privacy, Federated Learning, Functional Mean Estimation, Heterogeneous Privacy Constraints, Minimax Rates, Privacy-Accuracy Trade-off

1 Introduction

In the era of big data, sensitive information is often distributed across multiple sources in fields such as healthcare and financial services. Privacy concerns prevent direct data pooling, making it essential to develop efficient statistical inference methods that preserve privacy while leveraging the collective power of distributed data. Federated learning addresses this challenge by enabling organizations or groups to collaboratively train and improve a shared global machine learning model without sharing raw data externally, ensuring that privacy is maintained at each data source. Differential privacy (DP) [17] has emerged as the leading framework for providing

rigorous privacy guarantees, offering a principled approach to quantifying the extent to which individual privacy is protected within a dataset.

The problem of estimating the mean of random functions based on discretely sampled data arises naturally in functional data analysis and has significant applications in areas such as signal processing, biomedical imaging, environmental monitoring, and financial modeling [29, 19]. Classical methods have explored this problem extensively, with adaptive algorithms and theoretical results establishing optimal rates of convergence under varying sampling schemes [12, 23, 32, 30]. These methods emphasize a critical balance: increasing the number of subjects reduces variability across the population, while increasing the number of measurements per subject improves the precision of individual function estimates. This trade-off is fundamental to designing efficient sampling strategies.

Incorporating differential privacy into functional mean estimation introduces unique challenges. DP requires a trade-off between privacy and accuracy by adding noise to the data or the estimation process. For functional data, this trade-off is further complicated by the relationship between sampling intensity and privacy risk – taking more measurements per individual increases privacy risks, even though it enhances estimation accuracy. Developing DP-compliant estimators requires careful consideration of both privacy constraints and the intrinsic properties of functional data, diverging significantly from traditional non-private methods.

The federated learning paradigm introduces additional complexities to functional mean estimation under privacy constraints. In federated settings, data is distributed across multiple servers or entities, such as hospitals, mobile devices, or autonomous vehicles, each with distinct privacy requirements, see e.g. [8] and references therein for an overview. These settings often involve heterogeneous data characterized by varying numbers of individuals, measurements, and privacy budgets across servers. This heterogeneity exacerbates the challenges of balancing privacy and statistical accuracy.

This paper investigates the cost of privacy in functional mean estimation within the framework of Federated Differential Privacy (FDP). FDP extends DP principles to distributed environments, enabling privacy-preserving statistical analysis at various levels, including individual data holders, servers, or aggregated data. Private functional data analysis, especially when the private outputs are functions, has received recent attention, primarily within the central DP framework [21, 27]. The closest work to ours in the distributed private setting is [33], which studies private distributed estimation of a functional mean under a random-design model. Our framework differs in that we treat both independent and common-design regimes in a federated multi-server setting under heterogeneous privacy constraints, and in the overlapping heterogeneous independent-design setting our results yield a sharper rate characterization. We return to this comparison in Section 1.2; see also Section C for a detailed comparison.

Our work makes several key contributions. First, we establish minimax upper and lower bounds for the estimation error of the mean function under both common and independent design settings, highlighting key differences under FDP constraints. Second, we propose novel

algorithms that achieve the optimal trade-off between privacy and accuracy, accommodating heterogeneity in the number of individuals, measurements, and privacy budgets across servers. Third, we extend lower bound techniques by introducing innovative data processing methods that address information bottlenecks arising from server heterogeneity. Lastly, we quantify the fundamental limits of private functional mean estimation across various distributed environments, offering practical insights for privacy-preserving statistical analysis of functional data under diverse conditions, from local to central DP frameworks. We complement these theoretical and methodological results with an expanded simulation study and a real-data illustration in an emulated multi-server federated setting based on the Health and Retirement Study, illustrating the practical performance of the proposed methods under varying privacy budgets and heterogeneous server settings.

1.1 Problem Formulation

We consider the problem of distributed functional mean estimation under privacy constraints, where the data are partitioned across multiple servers. Specifically, for $s = 1, \ldots, S$, server s holds n_s independent observations.

For each $s = 1, \ldots, S$, let $X^{(s)}$ be a server-specific random function on $[0, 1]$, and let $X_1^{(s)}, \ldots, X_{n_s}^{(s)}$ be independent copies of $X^{(s)}$. We assume these server-specific laws share the same mean function

$$f(\cdot) = \mathbb{E}(X^{(s)}(\cdot)), \qquad s \in [S].$$

The goal is to estimate this common mean based on noisy observations from discrete locations on these curves distributed across S servers:

$$Y_{ij}^{(s)} = X_i^{(s)}(\zeta_{ij}^{(s)}) + \xi_{ij}^{(s)}, \quad i = 1, \ldots n_s \text{ and } j = 1, \ldots, m_s,$$

where $\zeta_{ij}^{(s)}$ are the sampling points and $\xi_{ij}^{(s)} \sim N(0, \sigma_s^2)$. We assume the common mean function f belongs to $\mathcal{H}^\alpha(R)$. Here, $\mathcal{H}^\alpha(R)$ denotes the Hölder ball of radius R within the class of α-Hölder continuous functions. In the classical setting without privacy constraints and distributed data, this problem has been extensively studied in the literature (e.g., [19, 12]). In this article, we focus on quantifying the cost of privacy in a distributed setup for this problem.

So far, we have not discussed the design of the experiment. We will consider two cases: the common design and the independent design. In the *common design* case, all individuals are observed on the same locations (grid points), i.e., for all $s \in [S]$ and $i \in [n_s]$, $\zeta_{ij}^{(s)} = \zeta_j$ for all $j \in [m_s]$. Note that in the common design case the design points are not considered to be private since they are shared among all individuals. In the *independent design* case, the grid points are random and distinct for each individual. In this case, the design points are considered to be private. This reflects real-world situations where individuals could be identified based on specific measurement time points. We make the simplifying assumption that they are i.i.d. draws from a uniform distribution on $[0, 1]$. This can be relaxed to non-uniform distributions that are,

e.g., bounded above and below by a constant, and the estimator presented in this paper can be easily adapted to non-uniform designs using the privacy-preserving technique presented in the Supplementary Material of [10]. The assumptions on the latent process X differ between these two settings: in the common design case, our results only require a pointwise subgaussian tail condition, while in the independent design case we assume the sample paths lie in $\mathcal{H}^\alpha(R)$ almost surely.

To account for privacy constraints, we adopt the general framework of distributed estimation under privacy constraints introduced in [10]. Let $Z^{(s)} = \{Z_i^{(s)}\}_{i=1}^{n_s}$ denote the dataset held by server s, with $Z_i^{(s)} = \{Y_{ij}^{(s)}, \zeta_{ij}^{(s)}\}_{j=1}^{m_s}$. Each server s outputs a (randomized) transcript $T^{(s)}$ based on $Z^{(s)}$, with the law of the transcript determined conditionally on $Z^{(s)}$, i.e., $\mathbb{P}(\cdot|Z^{(s)})$. The collection of transcripts $T = (T^{(s)})_{s=1}^S$ must satisfy $(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) = (\varepsilon_s, \delta_s)_{s=1}^S$-federated differential privacy (FDP), defined as follows:

Definition 1. The transcript $T = (T^{(s)})_{s=1}^S$ is $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$-federated differentially private (FDP) if for all $s \in [S]$, $A \in \mathscr{T}$ and $z, z' \in \mathcal{Z}^{n_s}$ differing in one individual datum it holds that

$$\mathbb{P}\left(T^{(s)} \in A | Z^{(s)} = z\right) \leq e^{\varepsilon_s} \mathbb{P}\left(T^{(s)} \in A | Z^{(s)} = z'\right) + \delta_s.$$

In the above definition,"differing in one datum" refers to being Hamming distance "neighbors." Specifically, local datasets $Z^{(s)}$ and $\tilde{Z}^{(s)}$ are *neighboring* if their Hamming distance is at most 1, calculated over $\mathcal{Z}^{n_s} \times \mathcal{Z}^{n_s}$. In other words, $\tilde{Z}^{(s)}$ can be derived from $Z^{(s)}$ by modifying at most one observation among $Z_1^{(s)}, \ldots, Z_{n_s}^{(s)}$. In the common design setup, where the design points are shared among all individuals, neighboring datasets share the same design points, and only the measurements Y differ. The smaller the values of ε_s and δ_s, the stricter the privacy constraint. We consider $\varepsilon_s \leq C_\varepsilon$ for $s = 1, \ldots, S$, with a fixed constant $C_\varepsilon > 0$ that does not affect the derived rates.

The Federated Differential Privacy (FDP) framework addresses scenarios where sensitive data is distributed across multiple parties, each generating outputs while maintaining differential privacy. In this distributed protocol, each server's transcript depends solely on its local data, with no direct communication or data exchange between servers. This framework is particularly relevant in settings like multi-site studies or trials conducted on the same population, where institutions (e.g., hospitals) aim to collaborate without sharing raw data due to privacy concerns. The FDP framework generalizes commonly studied settings, including the local DP model ($n_s = 1$), where privacy mechanisms operate at the individual level, and the central DP model ($S = 1$), where data is aggregated at a single location.

Each server transmits its transcript to a central server. Using all transcripts $T = (T^{(1)}, \ldots, T^{(S)})$, the central server computes an estimator $\hat{f} : \mathcal{T}^S \to \mathcal{F}$. We refer to the pair $(\hat{f}, \{\mathbb{P}(\cdot \mid Z^{(s)})\}_{s=1}^S)$ as a distributed estimation protocol. The class of distributed estimation protocols satisfying Definition 1 is denoted by $\mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$. We let $\mathbb{P}_f$ denote the joint law of transcripts and the $N = \sum_{s=1}^S n_s$ observations. We let $\mathbb{E}_f$ denote the expectation corresponding to $\mathbb{P}_f$.

The aim is to estimate the function f based on the distributed data. The difficulty of this estimation task arises from both the distributed nature of the data and privacy constraints that limit the sharing of information between servers. As in the conventional decision-theoretical framework, the estimation accuracy of a distributed estimator $\hat{f} \equiv \hat{f}(T)$ is measured by the integrated mean squared error (IMSE), $\mathbb{E}_f\|\hat{f} - f\|_2^2$, where the expectation is taken over the randomness in both the data and construction of the transcripts, and the quantity of particular interest is the *minimax risk* over the Hölder class $\mathcal{H}^\alpha(R)$ with $\alpha > 1/2$,

$$\inf_{\hat{f}\in\mathcal{M}(\boldsymbol{\varepsilon},\boldsymbol{\delta})} \sup_{f\in\mathcal{H}^\alpha(R)} \mathbb{E}_f\|\hat{f} - f\|_2^2. \tag{1}$$

This 'global' risk characterizes the difficulty of the federated learning problem over the parameter class when trying to infer the entire function underlying the data whilst adhering to the heterogeneous privacy constraints.

1.2 Related work

Federated Differential Privacy (FDP), as considered in this paper, applies DP at the level of local samples consisting of multiple observations. Optimal statistical inference under this framework has been studied in the context of nonparametric models for estimation [10], goodness-of-fit testing [9], and transfer learning for nonparametric classification [6].

The homogeneous setting has been explored for discrete distributions [26, 3] and parametric mean estimation [25, 28]. Additionally, [13] examines discrete distribution testing in a two-server setting ($m = 2$) with differing DP constraints. Like [10, 9], this paper works in the federated differential privacy framework where each server releases only a privatized local transcript. The main difference is the statistical task: [10] studies nonparametric regression, [9] studies nonparametric hypothesis testing, whereas here we study functional mean estimation from discretely sampled random curves, including both common and independent design. Complementary work on vertically partitioned private data includes [5] and [14] which studies estimation problems under server-level privacy constraints.

Upon completion of this work, we came across the paper [33], which considers a setting similar to ours and claims some partially overlapping results. There are important differences between the papers, however. Foremost, the upper and lower bounds in [33] exhibit a polynomial gap in the heterogeneous setting, whereas our bounds are tight up to polylogarithmic factors; see Section C in the Supplementary Material for a detailed comparison. Another important difference is that in [33], only the random design case is considered, which means the contrast with the fixed design setup are not treated in their work. Finally, the methods also differ: [33] relies on a sequential federated learning algorithm with multiple rounds of communication, while our procedure uses a single round in which each server transmits a transcript to a trusted central server. Part of this methodological difference is tied to the sampling-design assumptions: their random-design formulation allows a more general design distribution, whereas in our main presentation the design distribution is taken to be known (and uniform for simplicity).

In terms of lower bound techniques, several approaches have been specifically developed for private settings. [24] and [4] explore private adaptations of general methods, such as Fano, Assouad, and Le Cam techniques, for establishing lower bounds in the central DP setting. In contrast, [7] develops Van Trees-based lower bounds tailored to local DP, which do not directly extend to central DP. Similarly, the techniques in [1, 2] are designed for the local DP setting and lack straightforward extensions to central DP.

Building on these foundations, [10] extends the Van Trees-based lower bound techniques of [7] to the federated DP setting, addressing the complexities introduced by distributed data and federated privacy constraints. In this paper, we further advance these methodologies by applying them to the functional mean estimation problem. We introduce novel data processing techniques that account for the diverse information bottlenecks arising from user, measurement, and server heterogeneity, enabling the derivation of optimal lower bounds for heterogeneous configurations.

1.3 Organization of the paper

The rest of the paper is organized as follows. In Section 2, we present our main results, which summarize the fundamental performance limits of private functional mean estimation in the common and independent design settings. In Section 3, we describe the methods for differentially private functional mean estimation in the common and independent design settings. In Section 4, we present the lower bound for the common and independent design setting. Proofs of the main results are provided in the Supplementary Material.

1.4 Notation, definitions and assumptions

Throughout the paper, we shall write $N := \sum_{s=1}^{S} n_s$ and consider asymptotics in S, m_s, the n_s's and the privacy budget $(\boldsymbol{\epsilon}, \boldsymbol{\delta}) := \{\varepsilon_s, \delta_s\}_{s=1}^{S}$. Throughout the text we assume σ_s's are uniformly bounded below by a constant $\sigma_{\min}$. We assume that $N \to \infty$ and $\max_s \varepsilon_s = O(1)$. For two positive sequences a_k, b_k we write $a_k \lesssim b_k$ if the inequality $a_k \leq C b_k$ holds for some universal positive constant C. Similarly, we write $a_k \asymp b_k$ if $a_k \lesssim b_k$ and $b_k \lesssim a_k$ hold simultaneously and let $a_k \ll b_k$ denote that $a_k/b_k = o(1)$.

We use $a \vee b$ and $a \wedge b$ for the maximum and minimum, respectively, between a and b. For $k \in \mathbb{N}$, $[k]$ shall denote the set $\{1, \ldots, k\}$. Throughout the paper c and C denote universal constants whose value can differ from line to line. The Euclidean norm of a vector $v \in \mathbb{R}^d$ is denoted by $\|v\|_2$. For a matrix $M \in \mathbb{R}^{d \times d}$, the norm $M \mapsto \|M\|$ is the spectral norm and $\mathrm{Tr}(M)$ is its trace. Furthermore, we let I_d denote the $d \times d$ identity matrix. For a real-valued random variable Z, we define its Orlicz norm by $\|Z\|_{\psi_2} := \inf\left\{t > 0 : \mathbb{E}\exp(Z^2/t^2) \leq 2\right\}$.

2 Main results

In most practical settings, different servers often contain differing numbers of samples in terms of individuals, measurements per individual, and privacy parameters. We consider the most general setting where each server $s = 1, \ldots, S$ can have a different number of individuals n_s, a different number of measurements m_s for each individual in the server, and varying privacy parameters ε_s and δ_s. In the common design setting, our theory allows for heterogeneity in all of the aforementioned characteristics except for the design points.

Our main results are summarized in Theorem 1 and Theorem 2 below, which capture the fundamental performance limits of private functional mean estimation in the independent and common design settings, respectively.

2.1 Optimal performance when the design is independent

In this section, we consider the independent design setting. Theorem 1 below describes the minimax risk for differentially private estimators in the independent design setting. The rate-determining quantity in this setting is governed by the parameter $D_* \geq 1$:

$$D_*^{2\alpha} = \inf_{1 \leq D \leq D_*} \sum_{s=1}^{S} \min \left\{ D^{-1}\sigma_s^{-2} n_s m_s, D^{-2}\sigma_s^{-2} m_s n_s^2 \varepsilon_s^2, D^{2\alpha} n_s, D^{2\alpha-1} n_s^2 \varepsilon_s^2 \right\}. \tag{2}$$

Whenever the minimax risk is nontrivial, this equation admits a unique solution; see Section G of the Supplementary Material.

Theorem 1. *Assume that, for each server $s \in [S]$, the sample paths of $X^{(s)}$ belong to $\mathcal{H}^\alpha(R)$ almost surely and that*

$$\delta_s \log(1/\delta_s) \lesssim \left(\frac{n_s}{n_s \vee m_s} \varepsilon_s^2 D_*^{-1} \right) \wedge \left(\varepsilon_s^3 n_s^{3/4} D_*^{-3/2} \right),$$

where D_* solves (2). Then

$$\inf_{\hat{f} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{f \in \mathcal{H}^\alpha(R)} \mathbb{E}_f \|\hat{f} - f\|_2^2 \asymp \eta_N D_*^{-2\alpha},$$

where η_N is a polylogarithmic factor in $N = \sum_{s=1}^S n_s$ and $1/\delta_{\min}$, with $\delta_{\min} = \min_s \delta_s$.

The upper-bound part of Theorem 1 is established by Theorem 3, proved in Section D.1.2; the matching lower bound is Theorem 5, proved in Section F.1.

The optimization problem in (2) can be interpreted as the minimum of four terms, which correspond to the sampling error, the measurement error, and their respective private counterparts. Our upper bounds rely on solving the optimization problem (2), which determines, roughly speaking, the effective dimension of the problem. The effective dimension D_* balances the trade-offs between design complexity, sample sizes, noise levels, and the smoothness of the

underlying function class. More concretely, D_* is the largest effective resolution at which the aggregate information across servers still balances approximation bias. The four terms in (2) correspond to measurement information, privacy-limited measurement information, between-subject averaging information, and its privacy-limited counterpart. Thus, increasing m_s improves the measurement-related terms, increasing n_s improves both measurement and averaging terms, and increasing ε_s relaxes only the privacy-limited terms. The smoothness parameter α enters through the bias side $D^{2\alpha}$, determining how aggressively the effective resolution can increase before approximation error becomes dominant.

In the *non-private case*, where the privacy constraints are removed by setting $\varepsilon_s \to \infty$ for all $s \in [S]$, the rate simplifies to the classical non-private minimax rate. Notably, minimax optimal rates in the heterogeneous setup, where design points differ across samples, were not well understood prior to this work. The optimization problem in this case reveals how differences in sample sizes and design complexities across datasets interact with the smoothness of the underlying function to determine the rate.

2.2 Optimal performance when the design is common

In this section, we consider the case of common design, where $m_s = m$ for all $s = 1, \ldots, S$. For the main result of this section, we assume the common design points satisfy the quasi-uniform spacing condition

$$\zeta_1 \le \frac{C_\zeta}{m}, \qquad 1 - \zeta_m \le \frac{C_\zeta}{m}, \qquad \frac{c_\zeta}{m} \le \zeta_{j+1} - \zeta_j \le \frac{C_\zeta}{m}, \quad j = 1, \ldots, m-1, \tag{3}$$

for fixed constants $0 < c_\zeta \le C_\zeta < \infty$. This includes the canonical equispaced grid $\zeta_j = j/m$ as a special case. We note that our upper bounds for the common design setting hold for more general design point setups (see Section 3.2) and under the relaxed pathwise assumption

$$\sup_{t \in [0,1]} \|X_1^{(s)}(t)\|_{\psi_2} \le C_X \sigma_s \quad \text{for all } s \in [S] \tag{4}$$

for some fixed constant $C_X > 0$.

Theorem 2 below describes the minimax risk for differentially private estimators in the common design setting. The rate-determining quantity in this setting is governed by the parameter $D_* \ge 1$, which is the solution to the following equation:

$$D_*^{2\alpha} = m^{2\alpha} \bigwedge \sum_{s=1}^{S} \min\left\{\sigma_s^{-2} n_s, D_*^{-1} \sigma_s^{-2} n_s^2 \varepsilon_s^2\right\}. \tag{5}$$

Whenever the minimax risk is nontrivial, this equation admits a unique solution; see Section G of the Supplementary Material.

Theorem 2. *Assume that the common design satisfies* (3) *and that* (4) *holds. Let D_* be the solution to* (5)*. Further suppose that, for each $s \in [S]$, $\delta_s \log(1/\delta_s) \lesssim D_*^{-1} \varepsilon_s^2$. Then*

$$\inf_{\hat{f} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{f \in \mathcal{H}^\alpha(R)} \mathbb{E}_f \|\hat{f} - f\|_2^2 \asymp \eta_N D_*^{-2\alpha},$$

where η_N is a polylogarithmic factor in $N = \sum_{s=1}^{S} n_s$ and $1/\delta_{\min}$, with $\delta_{\min} = \min_s \delta_s$.

The upper-bound part of Theorem 2 is established by Theorem 4, proved in Section E.1; the matching lower bound is Theorem 6, proved in Section F.2.

It is instructive to consider two specific limiting cases to better understand the implications of this result. First, in the *non-private case*, where the privacy constraints are removed by setting $\varepsilon_s \to \infty$ for all $s \in [S]$, the rate simplifies to the known minimax rates established in [12]. This scenario corresponds to the classical non-private setting where privacy considerations do not impose any restrictions. The effective rate is determined solely by the design complexity and sample sizes, without the additional penalties introduced by privacy constraints.

Second, in the *large-sample limit*, where $n_s \to \infty$ for all $s \in [S]$, the rate reduces to $m^{-2\alpha}$. Here, the noise becomes negligible due to the abundance of samples, and the discretization error from the common design dominates the estimation error. This reflects a regime where the design complexity m dictates the achievable performance, independent of the privacy constraints or noise levels.

In order to gain more fine-grained insight it is helpful to look at the homogeneous setting, which we do in Section 2.3.

2.3 Comparison of the common and independent design settings

In this section, we compare the minimax rates of convergence for the common and independent design settings under the homogeneous case, where $n_s = n$, $m_s = m$, $\varepsilon_s = \varepsilon$, $\sigma_s = 1$ and $\delta_s = \delta$. While in practice, settings are more often heterogeneous than not, the homogenous case allows for a more straightforward formulation of the minimax rate than the implicit ones of Theorems 1 and 2. In this simplified formulation, the rates between the independent and common designs are more easily compared. The comparison highlights the fundamental differences in statistical efficiency and the impact of differential privacy constraints across these two design settings.

We begin by stating the minimax risks for the homogeneous setup under the two design settings, starting with the independent design setting.

Corollary 1. *Under the same conditions as in Theorem 1 it holds that*

$$\inf_{\hat{f} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{f \in \mathcal{H}^\alpha(R)} \mathbb{E}_f \|\hat{f} - f\|_2^2 \asymp \eta_N \left((Sn)^{-1} + (Smn)^{-\frac{2\alpha}{2\alpha+1}} + (Smn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+2}} + (Sn^2\varepsilon^2)^{-1} \right), \quad (6)$$

where η_N is a polylogarithmic factor in $N = Sn$ and $1/\delta$.

For the common design setting, the minimax rate of convergence in the homogeneous setup is given by the following corollary.

Corollary 2. *Under the same conditions as Theorem 2 it holds that*

$$\inf_{\hat{f} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{f \in \mathcal{H}^\alpha(R)} \mathbb{E}_f \|\hat{f} - f\|_2^2 \asymp \eta_N \left((Sn)^{-1} + m^{-2\alpha} + (Sn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+1}} \right), \quad (7)$$

where η_N is a polylogarithmic factor in $N = Sn$ and $1/\delta$.

In the common design case, the rate of convergence in (7) is dominated by the privacy-related term $(Sn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+1}}$ when $\varepsilon \lesssim \min\left(S^{-1/2}n^{-1}m^{\alpha+1/2}, S^{1/4\alpha}n^{(1-2\alpha)/4\alpha}\right)$. In contrast, under the independent design setting, the rate is given by the combination $(Smn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+2}} + (Sn^2\varepsilon^2)^{-1}$ when privacy constraints dominate.

If the rate is dominated by the second term $(Sn^2\varepsilon^2)^{-1}$, it is strictly faster than the common design privacy rate. If the first term $(Smn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+2}}$ dominates, it is faster than the common design rate whenever

$$(Smn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+2}} \lesssim (Sn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+1}} \quad \text{iff} \quad \varepsilon \lesssim S^{-1/2}n^{-1}m^{\alpha+1/2}.$$

Thus, whenever the "privacy-related" terms dominate, an independent design achieves a faster rate of convergence, meaning the cost of privacy is consistently lower in this setting. Intuitively, under common design additional measurements are taken on the same shared grid, so they mainly increase the amount of information that must be privatized, whereas under independent design they also improve the random coverage of the domain.

When privacy constraints are binding, the independent-design setting benefits from having both more individuals and more measurements per individual, whereas under common design increasing the number of measurements per individual offers little benefit. Although this may seem counterintuitive, since the design points are private only in the independent-design setting, the difference comes from how additional measurements affect effective information. Under common design, increasing m mainly adds shared information that must be privatized, without substantially enriching the design. Under independent design, by contrast, increasing m also improves random coverage of the domain, so the statistical gain can continue to outweigh the added privacy cost.

3 Methods for Differentially Private Functional Mean Estimation

We now describe the methods for differentially private functional mean estimation in the independent and common design settings. As revealed by Theorems 1 and 2 , these settings exhibit distinct principal phenomena and require different approaches. We present an optimal algorithm for each setting, both of which provably achieve the minimax rate of convergence. The theorems accompanying the algorithms in this section provide the upper bounds for the respective theorems mentioned earlier, starting with the independent design case.

3.1 An optimal algorithm for the independent design setting

In the independent design setting, we consider a scenario where all design points are distinct and treated as private information. We additionally assume that for each server $s \in [S]$, $X_i^{(s)} \in \mathcal{H}^\alpha(R)$ for $i \in [n_s]$. The challenge in this setup, compared to e.g. private nonparametric

regression as studied in [10], stems from the fact that the model itself is effectively subject to three different sources of noise: the per-subject $X_i^{(s)}$-function noise, the noise stemming from the randomness in the design points, and the measurement noise. An optimal privacy mechanism is calibrated to account for all three sources of noise, ensuring that the privacy budget for each server is allocated efficiently across the different components. This means that the procedure must differ from server to server, depending on the number of individuals, measurements per individual, and privacy parameters.

Our approach is based on a private, truncated projection estimator, where we project onto compactly supported orthonormal basis functions. This approach allows effective privatization of both the design points and the outcome values. As basis functions, we use $A > \alpha$-regular wavelets, which form an orthonormal basis of $L_2[0,1]$. Wavelet bases allow characterization of Hölder spaces, with α capturing the decay of wavelet coefficients. For further details, see e.g. [22].

For any $A \in \mathbb{N}$ one can follow Daubechies' construction of the father $\phi(\cdot)$ and mother $\psi(\cdot)$ wavelets with A vanishing moments and bounded support on $[0, 2A-1]$ and $[-A+1, A]$, respectively, for which we refer to [16]. The basis functions are then obtained as

$$\big\{\phi_{l_0+1,m}, \psi_{lk} : m \in \{0, ..., 2^{l_0+1}-1\}, \quad l \geq l_0+1, \quad k \in \{0, ..., 2^l-1\}\big\},$$

with $\psi_{lk}(x) = 2^{l/2}\psi(2^l x - k)$, for $k \in [A-1, 2^l - A]$, and $\phi_{l_0+1,m}(x) = 2^{\frac{l_0+1}{2}}\phi(2^{l_0+1}x - m)$, for $m \in [0, 2^{l_0+1} - 2A]$, while for other values of k and m, the functions are specially constructed to form a basis with the required smoothness property. For notational convenience, we write $\psi_{l_0,k} := \phi_{l_0+1,k}$ for $k = 0, \ldots, 2^{l_0+1}-1$. A function $f \in L_2[0,1]$ can then be expressed as

$$f = \sum_{l=l_0}^{\infty} \sum_{k=0}^{2^l-1} f_{lk}\psi_{lk},$$

where $f_{lk} = \int f\psi_{lk}$. A well-known estimator for the wavelet coefficients f_{lk}, in the absence of privacy constraints is given by

$$\widehat{f}_{lk}^{(s)} = \frac{1}{n_s}\sum_{i=1}^{n_s} \frac{1}{m_s}\sum_{j=1}^{m_s} Y_{ij}^{(s)}\psi_{lk}\left(\zeta_{ij}^{(s)}\right),$$

where $Y_{ij}^{(s)}$ are the observations and $\zeta_{ij}^{(s)}$ are the corresponding design points.

For notational simplicity, let us define the contribution of a single individual $i \in [n_s]$, on server $s \in [S]$ to the estimator as:

$$U_{i,lk}^{(s)} = \frac{1}{m_s}\sum_{j=1}^{m_s} Y_{ij}^{(s)}\psi_{lk}(\zeta_{ij}^{(s)}). \tag{8}$$

Thus, the standard (unclipped) estimator can be rewritten as: $\widehat{f}_{lk}^{(s)} = \frac{1}{n_s}\sum_{i=1}^{n_s} U_{i,lk}^{(s)}$. In order to make an informed decision on how to clip, we require sharp concentration bounds on $U_{i,lk}^{(s)}$.

The following lemma establishes the concentration of $U_{i,lk}^{(s)}$ for a fixed l and k, using Bernstein's inequality.

Lemma 1. *For a fixed l and k, $\{U_{i,lk}^{(s)}\}_{i\in[n_s]}$ as defined in (8) satisfies*

$$\mathbb{P}_f\Big(\forall i,\ |U_{i,lk}^{(s)}| < \tau_l^{(s)}\Big) \geq 1-\gamma$$

with

$$\tau_l^{(s)} = 2\sigma_s C'(2c\log N)^{3/2}\left(\sqrt{\frac{1}{m_s}} + \frac{1}{3}\|\psi\|_\infty \frac{2^{l/2}}{m_s}\right) + R2^{-l(\alpha+1/2)} \quad \textit{and} \quad \gamma = (N)^{-c}. \tag{9}$$

where C' depends on c, $\sigma_{\min}$, and $\|\psi\|_\infty$.

A proof of the lemma is given in Section D.1.2 of the Supplementary Material. We define what is effectively the local clipped estimator for the l,k-th wavelet coefficient as

$$\widehat{f}_{lk}^{\tau,s} = \frac{1}{n_s}\sum_{i=1}^{n_s}\left[U_{i,lk}^{(s)}\right]_{\tau_l^{(s)}}, \tag{10}$$

where $\left[U_{i,lk}^{(s)}\right]_{\tau_l^{(s)}}$ denotes the clipping of $U_{i,lk}^{(s)}$ to the range $[-\tau_l^{(s)}, \tau_l^{(s)}]$. Clipping at this magnitude ensures that the estimator has bounded sensitivity, enabling it to be privatized effectively.

To achieve differential privacy, we aim to employ the Gaussian mechanism. However, adding Gaussian noise to the clipped averages of (10) directly does not balance the amount of signal versus sensitivity present in each wavelet coefficient. Instead, we employ a modified Gaussian mechanism, akin to ideas used in anisotropic private mean estimation; see, for example, [15].

The coordinates of $\widehat{f}_{lk}^{\tau}$ exhibit greatly varying sensitivities across the resolution levels, proportional to $\tau_l^{(s)}$. In order to obtain the right signal-sensitivity trade-off, we rescale each coordinate, yielding at the l-th resolution level:

$$\tilde{f}_{lk}^{\tau,s} = (\tau_l^{(s)}\sqrt{2^l \wedge m_s})^{-1}\widehat{f}_{lk}^{\tau,s}.$$

The ℓ_2 sensitivity of the vector $\{\tilde{f}_{lk}^{\tau,s}, k = 0,\dots,2^l-1, l_0 \leq l \leq L\}$ is bounded as shown in Lemma 2.

Lemma 2. *Let $Z^{(s)}$ and $\tilde{Z}^{(s)}$ be any neighboring datasets. Let $\tilde{f}^{\tau,s}(Z^{(s)}), \tilde{f}^{\tau,s}(\tilde{Z}^{(s)})$ denote the vectors of coefficient estimators $\{\tilde{f}_{lk}^{\tau,s}, k = 0,\dots,2^l-1, l_0 \leq l \leq L\}$ computed on the basis of $Z^{(s)}$ and $\tilde{Z}^{(s)}$ respectively. Then:*

$$\left\|\tilde{f}^{\tau,s}(Z^{(j)}) - \tilde{f}^{\tau,s}(\tilde{Z}^{(j)})\right\|_2 \leq c_A\frac{\sqrt{L}}{n_s},$$

where c_A is a computable constant depending only on the choice of wavelet basis.

Using the Gaussian mechanism (for details, see Appendix A of [18]), we obtain a $(\varepsilon_s, \delta_s)$-differentially private estimator

$$\tilde{f}_{lk}^{P,s} = \tilde{f}_{lk}^{\tau,s} + \widetilde{W}_{lk}^{(s)}, \quad \widetilde{W}_{lk}^{(s)} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \frac{4c_A^2 L \log(2/\delta_s)}{n_s^2 \varepsilon_s^2}\right).$$

To construct the server-specific private estimator $\widehat{f}_{lk}^{P,s}$, we rescale $\tilde{f}_{lk}^{P,s}$ back to the original scale, yielding

$$\widehat{f}_{lk}^{P,s} = \widehat{f}_{lk}^{\tau,s} + W_{lk}^{(s)}, \quad W_{lk}^{(s)} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \frac{4c_A^2 L (2^l \wedge m_s)(\tau_l^{(s)})^2 \log(2/\delta_s)}{n_s^2 \varepsilon_s^2}\right).$$

Each server s then outputs a transcript

$$T_L^{(s)} = \{\widehat{f}_{lk}^{P,s} : k = 0, \ldots, 2^l - 1, l = l_0, \ldots, L\}.$$

The final estimator $\widehat{f}^P$ is obtained via a post-processing step, where the transcripts are aggregated using weights that account for the heterogeneity of the servers. These weights depend on the resolution level l, the number of local observations n_s, the number of design points m_s, the local privacy constraints ε_s, and the smoothness level α. Specifically, the final estimator is:

$$\widehat{f}_L^P = \sum_{l=l_0}^{L} \sum_{k=0}^{2^l-1} \widehat{f}_{lk}^P \psi_{lk} \quad \text{where} \quad \widehat{f}_{lk}^P = \sum_{s=1}^{S} w_l^{(s)} \widehat{f}_{lk}^{P,s} \tag{11}$$

with $\sum_{s=1}^S w_l^{(s)} = 1$ for all $l \geq l_0$. The weights $w_l^{(s)}$ are chosen according to variance-proxy quantities for $\widehat{f}_{lk}^{P,s}$:

$$w_l^{(s)} = \frac{u_l^{(s)}}{\sum_{s=1}^S u_l^{(s)}}, \quad \text{where} \quad u_l^{(s)} = \sigma_s^{-2} n_s m_s \wedge n_s 2^{l(2\alpha+1)} \wedge \sigma_s^{-2} 2^{-l} m_s n_s^2 \varepsilon_s^2 \wedge 2^{2l\alpha} n_s^2 \varepsilon_s^2. \tag{12}$$

Remark 1. The weights $w_l^{(s)}$ account for the heterogeneity in data distribution across servers, including variations in the number of observations (n_s), the number of design points (m_s),the variance (σ_s, and the privacy budgets (ε_s). By being inversely proportional to the variance of $\widehat{f}_{lk}^{P,s}$, the weights prioritize servers that contribute more reliable estimates. The structure of $u_l^{(s)}$ incorporates terms that reflect the impact of resolution level l, smoothness parameter α, and privacy constraints.

We summarize our final estimator in Algorithm 1.

Theorem 3. *Assume for each server $s \in [S]$, $X_i^{(s)} \in \mathcal{H}^\alpha(R)$ for $i \in [n_s]$. Let $\widehat{f}_L^P$ be the estimator defined in Algorithm 1. Let D_* satisfy* (2)*, and choose $L = \lfloor \log_2 D_* \rfloor$. Then*

$$\sup_{f \in \mathcal{H}^\alpha(R)} \mathbb{E}_f \|\widehat{f}_L^P - f\|_2^2 \lesssim (\log N)^5 \log(2/\delta_{\min}) (D_*)^{-2\alpha}.$$

where $\delta_{\min} = \min_s \delta_s$.

Algorithm 1 Private Estimation via Truncated Projection in Independent Design

1: **Input:** Observations $Y_{ij}^{(s)}$, design points $\zeta_{ij}^{(s)}$, wavelet basis ψ_{lk}, resolution level L, privacy budgets ε_s, δ_s

2: **Output:** Final estimator $\widehat{f}_L^P$

3: **Clipped Coefficients:** Each server computes:

$$\widehat{f}_{lk}^{\tau,s} = \frac{1}{n_s}\sum_{i=1}^{n_s}\left[\frac{1}{m_s}\sum_{j=1}^{m_s} Y_{ij}^{(s)}\psi_{lk}(\zeta_{ij}^{(s)})\right]_{\tau_l^{(s)}} \quad \forall\, l = l_0,\ldots,L,\ k = 0,\ldots,2^l-1,$$

where $\tau_l^{(s)}$ is given by (9).

4: **Privatized Coefficients:** Apply the modified Gaussian mechanism:

$$\widehat{f}_{lk}^{P,s} = \widehat{f}_{lk}^{\tau,s} + W_{lk}^{(s)}, \quad W_{lk}^{(s)} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \frac{4c_A^2 L(2^l \wedge m_s)(\tau_l^{(s)})^2 \log(2/\delta_s)}{n_s^2\varepsilon_s^2}\right).$$

5: **Weighted Aggregation:** Aggregate across servers:

$$\widehat{f}_{lk}^{P} = \sum_{s=1}^{S} w_l^{(s)}\widehat{f}_{lk}^{P,s}, \quad w_l^{(s)} = \frac{u_l^{(s)}}{\sum_{s=1}^{S} u_l^{(s)}},\ u_l^{(s)} = \sigma_s^{-2} n_s m_s \wedge n_s 2^{l(2\alpha+1)} \wedge \sigma_s^{-2} 2^{-l} m_s n_s^2\varepsilon_s^2 \wedge 2^{2l\alpha} n_s^2\varepsilon_s^2.$$

6: **Final Estimator:** Combine wavelet coefficients:

$$\widehat{f}_L^P(x) = \sum_{l=l_0}^{L}\sum_{k=0}^{2^l-1} \widehat{f}_{lk}^P \psi_{lk}(x).$$

A proof is given in Section D.1.2.

Remark 2. The parameter D_*, which satisfies the rate-determining relation (2), can be interpreted as the effective dimension of the problem. It reflects the interplay between the resolution level L and the sample size $N = \sum_{s=1}^{S} n_s$. Larger D_* corresponds to higher resolution levels and greater flexibility in representing the function f, while also increasing the complexity of the estimation task. The optimal choice of D_* balances this trade-off to achieve the minimax rate.

Remark 3. In the absence of privacy constraints ($\varepsilon_s \to \infty$), the result in Theorem 3 obtains the minimax rates for the setting where different individuals may have different numbers of design points m_s. To the best of our knowledge, these rates were not previously established in the literature.

3.2 An optimal algorithm for the common design setting

In the common design setup, the design points $\zeta_1, \zeta_2, \ldots, \zeta_m$ are shared among all individuals and servers. These points are assumed to be publicly known, requiring no privatization. Here, the challenge lies in the fact that the design points are shared across all servers, which means that an optimal private estimator must balance the discretization error due to the shared design points with the privacy leakage.

Our estimation strategy follows three main steps:

(a) First, we obtain privatized averages of observations at each design point, effectively creating private, noisy estimates of the function at these specific points.

(b) Second, if the number of design points is large, we introduce a bagging step that divides the design points into groups.

(c) Third, these privatized averages present an inverse problem at the central server, which we solve using local polynomial regression. This regression effectively interpolates across the design points.

Private averages at the design points of first step effectively pose an inverse problem at the central server; solved by local polynomial regression in the third step. The noise in this inverse problem at the central server does not only depend on measurement noise, but also on the noise added by the privacy mechanism. If the number of design points is large, the noise added by the privacy mechanism needs to increase to avoid the privacy leakage due to revealing more measurements per individual.

The second step employing bagging mitigates this by dividing the design points into groups and computes smoothed estimators that interpolate across groups. Intuitively, the purpose of the bagging step is used to effectively reduce the number of measurements m, in the scenario where the measurements reveal too much information about the individuals if they were all communicated to the central server. In terms of attaining the minimax rate, the bagging step is superfluous, in the sense that one could also effectively use a smaller number of measurements by taking a selection of design points. Bagging, however, is a more practical approach, as it a straightforward implementation that is more robust to the choice of design points.

Next, we describe the algorithm for the common design formally.

Privatized averages at the design points Given S servers and a set of observations from a function f at deterministic points $\zeta_1, \zeta_2, \ldots, \zeta_m$, the privatized mean $\bar{Y}_j^{P,s}$ at design point ζ_j on server s is computed as:

$$\bar{Y}_j^{P,s} = \frac{1}{n_s} \sum_{i=1}^{n_s} \left[Y_{i,j}^{(s)} \right]_{\tau_s} + W_j^{(s)},$$

where $\left[Y_{i,j}^{(s)}\right]_{\tau_s}$ represents the clipped observations with clipping threshold $\tau_s = M\sigma_s\sqrt{\log N}$ (M is suitably chosen), and $W_j^{(s)} \sim \mathcal{N}\left(0, \frac{4\tau_s^2 m \log(2/\delta_s)}{n_s^2\varepsilon_s^2}\right)$ is a Gaussian noise term ensuring differential privacy. Here, N is the total sample size across all servers, n_s is the sample size on server s, and ε_s is the privacy budget for server s.

Each server outputs a transcript $T^{(s)} = (\bar{Y}_j^{P,s})_{j\in[m]}$. The central server aggregates these transcripts using inverse variance weighting:

$$\bar{Y}_j^P = \sum_{s=1}^{S} w_s \bar{Y}_j^{P,s},$$

where the weights w_s are designed to account for server heterogeneity:

$$w_s = \frac{u_s}{\sum_{s=1}^{S} u_s}, \quad u_s = \sigma_s^{-2}\left(D_*^{-1} n_s^2\varepsilon_s^2 \wedge n_s\right).$$

where D_* solves (5).

Bagging Let

$$m_0 = \lfloor D_* \rfloor, \qquad B = \left\lfloor \frac{m}{m_0} \right\rfloor,$$

where D_* solves (5). We divide the design-point indices into B disjoint groups $\{G_b\}_{b\in[B]}$, each of size exactly m_0. In the equispaced case $\zeta_j = j/m$, one convenient choice is

$$G_b = \{\, B(j-1) + b \in [m] : j = 1, \ldots, m_0 \,\}, \qquad b \in [B].$$

The corresponding design points in group b are then $(\zeta_j)_{j\in G_b}$.

For each group G_b, a local polynomial estimator $\widehat{f}_b^P(x)$ is computed based on the privatized means $\bar{Y}_j^P$ corresponding to the design points in G_b. Finally, the overall estimator $\widehat{f}_B^P(x)$ is obtained by averaging across all groups:

$$\widehat{f}_B^P(x) = \frac{1}{B}\sum_{b=1}^{B} \widehat{f}_b^P(x).$$

Function approximation via local polynomials The local polynomial estimator approximates the function $f \in \mathcal{H}^\alpha(R)$ by expanding it around x as:

$$f(z) = f(x) + f'(x)(z-x) + \cdots + \frac{f^{(p)}(x)}{p!}(z-x)^p,$$

where $p = \lfloor \alpha \rfloor$. This expansion can be written compactly as:

$$f(z) = \Theta^\top(x) V\left(\frac{z-x}{h}\right),$$

where $V(u) = \left(1, u, \frac{u^2}{2!}, \ldots, \frac{u^p}{p!}\right)^\top$ is the vector of scaled monomials, and $\Theta(x) = \left(f(x), f'(x)h, \ldots, f^{(p)}(x)h^p\right)^\top$ represents the function's coefficients at x, scaled by powers of the bandwidth h.

To estimate $f(x)$, we solve a weighted least squares problem for each group:

$$\widehat{\Theta}_b(x) = \arg\min_{\Theta \in \mathbb{R}^{p+1}} \sum_{j \in G_b} \left[\bar{Y}_j^P - \Theta^\top V\left(\frac{\zeta_j - x}{h}\right)\right]^2 K\left(\frac{\zeta_j - x}{h}\right).$$

where K can be any smooth enough kernel function satisfying

$$\mathrm{supp}(K) \subseteq [-1, 1], \quad \|K\|_\infty < \infty, \quad \text{and} \quad \int K(u)\, du = 1.$$

The first component of $\widehat{\Theta}_b(x)$, corresponding to $\hat{f}_b^P(x)$, is the estimate of $f(x)$ for group b.

The local polynomial estimator $\hat{f}_b^P(x)$ for the b-th group can be shown to be equal to:

$$\hat{f}_b^P(x) = \sum_{j \in G_b} \bar{Y}_j^P W_{b,j}^*(x),$$

where the weights $W_{b,j}^*(x)$ are determined by the kernel and the design points of the bth group:

$$W_{b,j}^*(x) = \frac{1}{m_0 h} V^\top(0) B_{b,x}^{-1} V\left(\frac{\zeta_j - x}{h}\right) K\left(\frac{\zeta_j - x}{h}\right). \tag{13}$$

The matrix $B_{b,x}$ is the "effective" weighted design matrix for group b induced by K, given by

$$B_{b,x} = \frac{1}{m_0 h} \sum_{j \in G_b} V\left(\frac{\zeta_j - x}{h}\right) V\left(\frac{\zeta_j - x}{h}\right)^\top K\left(\frac{\zeta_j - x}{h}\right).$$

For the subsequent analysis, we rely on the following assumptions, where the choice of kernel and groups allows for additional flexibility. We remark that both assumptions (a) and (b) hold in particular for regular equi-spaced design under mild conditions on the Kernel K. See for example Lemma 1.5 in [31].

(a) $\lambda_{\min}(B_{b,x}) \geq \lambda_0$ for all $b \in [B]$, and $x \in [0, 1]$.

(b) For $b \in [B]$, for some constant a, and any measurable set A, we have

$$\frac{1}{m_0} \sum_{j \in G_b} \mathbb{I}(\zeta_j \in A) \leq a \max\left(\mathrm{Leb}(A), \frac{1}{m_0}\right).$$

Assumption (a) ensures that the local design points in each group provide sufficient information for local polynomial estimation. Together with (b), these assumptions ensure that within each group the design points $(\zeta_j)_{j \in G_b}$ are sufficiently spread over the interval $[0, 1]$.

Algorithm 2 summarizes the differentially private protocol. By averaging across all groups in (14), the bagging estimator as $\hat{f}_B^P(x)$ is computed in the central server. The following theorem establishes the convergence rate of this estimator.

Algorithm 2 Private Estimation via Smoothing and Bagging for Common Design

1: **Input:** Design points $\zeta_1, \ldots, \zeta_m$, S servers, observations $Y_{i,j}^{(s)}$, privacy budgets ε_s, groups $\{G_b\}_{b\in[B]}$ each of size m_0 , bandwidth h, kernel K

2: **Output:** Final estimator $\hat{f}_B^P(x)$

3: **Privatized Means:** Each server computes privatized means:

$$\bar{Y}_j^{P,s} = \frac{1}{n_s}\sum_{i=1}^{n_s}\left[Y_{i,j}^{(s)}\right]_{\tau_s} + W_j^{(s)}, \quad W_j^{(s)} \sim \mathcal{N}\left(0, \frac{4\tau_s^2 m \log(2/\delta_s)}{n_s^2\varepsilon_s^2}\right),$$

where $\tau_s = M\sigma_s\sqrt{\log N}$.

4: **Aggregate Means:** Central server aggregates:

$$\bar{Y}_j^P = \sum_{s=1}^{S} w_s \bar{Y}_j^{P,s}, \quad w_s = \frac{u_s}{\sum_{s=1}^{S} u_s}, \quad u_s = \sigma_s^{-2}\left(m_0^{-1}n_s^2\varepsilon_s^2 \wedge n_s\right).$$

5: **Group-wise Estimation:** For each group G_b, compute local polynomial estimator:

$$\widehat{f}_b^P(x) = \sum_{j\in G_b} \bar{Y}_j^P W_{b,j}^*(x),$$

where $W_{b,j}^*(x)$ are kernel-based weights given by (13).

6: **Final Estimator:** Aggregate across groups:

$$\widehat{f}_B^P(x) = \frac{1}{B}\sum_{b=1}^{B}\widehat{f}_b^P(x). \tag{14}$$

Theorem 4. *Assume that the sample paths $X_1^{(s)}, s \in [S]$ satisfy (4). Let $\widehat{f}_B^P$ be the estimator defined in Algorithm 2. Let D_* solve (5), set $m_0 = \lfloor D_* \rfloor$, choose groups $(G_b)_{b\in[B]}$ each of size m_0, and set the bandwidth h to be $1/m_0$. We also assume that for each group $(G_b)_{b\in[B]}$, the design points $(\zeta_j)_{j\in G_b}$ satisfy assumptions (a) and (b). Then,*

$$\sup_{f\in\mathcal{H}^\alpha(R)} \mathbb{E}_f\|\hat{f} - f\|_2^2 \lesssim \log(N)\log(2/\delta_{\min})\left(D_*^{-2\alpha} \vee m^{-2\alpha}\right),$$

where D_* solves (5) and $\delta_{\min} = \min_s \delta_s$.

A proof is given in Section E.1.

Remark 4. The same minimax risk as in Theorem 4 can be achieved by using data from only one of the groups of size m_0 instead of aggregating across all B groups. This is because, when using a single group, the amount of noise added to ensure differential privacy scales with m_0, which is smaller than or equal to m. Consequently, the effective privacy noise is reduced, and the overall estimation error remains comparable. While using all B groups can improve practical

stability/performance by averaging across groups, it does not offer a theoretical advantage in terms of the minimax rate.

Remark 5 (Clipping under stronger pathwise assumptions). If one instead has $X_i^{(s)} \in \mathcal{H}^\alpha(R)$, as opposed to (4) then $\forall i, s\ \|X_i^{(s)}\|_\infty \leq C_{\alpha,R}$, and one may use $\tau_s = \sigma_s\sqrt{2\log N} + C_{\alpha,R}$. This changes only the clipping constant, not the convergence rate.

4 Lower bounds

As a complement to the upper bounds, we provide lower bounds, which together establish the minimax rates for the federated learning problem; resulting in Theorems 2 and 1. We first present the lower bound for the independent design setting, followed by the lower bound for the common design setting.

4.1 Lower bound for independent design

We first present the lower bound, before discussing the proof.

Theorem 5. *Suppose that, for every $s \in [S]$, the sample paths of $X^{(s)}$ belong to $\mathcal{H}^\alpha(R)$ almost surely and $\delta_s \log(1/\delta_s) \lesssim \left(\frac{n_s}{n_s \vee m_s}\varepsilon_s^2 D_*^{-1}\right) \wedge \left(\varepsilon_s^3 n_s^{3/4} D_*^{-3/2}\right)$. Let D_* solve (2). Then,*

$$\inf_{\hat{f} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{f \in \mathcal{H}^\alpha(R)} \mathbb{E}_f \|\hat{f} - f\|_2^2 \gtrsim D_*^{-2\alpha}. \tag{15}$$

The full proof is given in Section F.1.

Below, we briefly sketch the proof of Theorem 5. The challenge in proving the lower bound stems from the heterogeneity of the servers. Essentially, for each respective server, either the sample size, the number of design points, or the privacy constraints can form the bottleneck.

The proof involves a careful construction of a data generating process that allows us to capture these different bottlenecks for each server. From a technical standpoint, the heterogeneous nature of the servers is captured by considering the Fisher information of the model in an appropriate finite-dimensional sub-model. The Fisher information tensorizes along the servers, which allows us to provide separate information-theoretic arguments for each server.

To provide a more detailed description of the argument, we introduce the aforementioned notions formally. Consider the 2^L-dimensional sub-model given by

$$\mathcal{H}^\alpha_{R,L} = \left\{ f \in \mathcal{H}^\alpha(R) \,:\, f = \sum_{k=1}^{2^L} f_k \phi_{Lk},\ f_k \in [-2^{-L(\alpha+1/2)-1}R, 2^{-L(\alpha+1/2)-1}R] \right\}, \tag{16}$$

where ϕ_{Lk} are the wavelet basis functions at resolution level L. Let μ and ν denote dominating measures for $\mathbb{P}_f^{Y^{(s)}}$ and $\mathbb{P}_f^{(Y^{(s)}X^{(s)})}$, respectively, for $f \in \mathcal{H}^\alpha_{R,L}$. Define $f_L = (f_k)_{k=1}^{2^L}$ to be the

vector of coefficients of f in the wavelet basis. The transcript $T^{(s)}$ induces a "Fisher information", which can be computed to be equal

$$I_f^{Y^{(s)}|T^{(s)}} = \mathbb{E}_f \, \mathbb{E}_f \left[S_f^{Y^{(s)}} \middle| T^{(s)} \right] \mathbb{E}_f \left[S_f^{Y^{(s)}} \middle| T^{(s)} \right]^\top \tag{17}$$

where $S_f^{Y^{(s)}}$ is given by

$$S_f^{Y_i^{(s)}} = \sigma_s^{-2} \left(\sum_{j=1}^{m_s} \left(Y_{ij}^{(s)} - \sum_{k=1}^{2^L} f_k \phi_{Lk}(\zeta_{ij}^{(s)}) \right) \phi_{Ll}(\zeta_{ij}^{(s)}) \right)_{l \in [2^L]}, \tag{18}$$

which can be seen as a "score function" of the observational model for $Y^{(s)}$ in the sub-model $\mathcal{H}_{R,L}^{\alpha}$. Let $I_f^{Y^{(s)}}$ denote the Fisher information of the observational model for $Y^{(s)}$ in the sub-model $\mathcal{H}_{R,L}^{\alpha}$, which equals $\mathbb{E}_f S_f^{Y^{(s)}} S_f^{Y^{(s)}\top}$.

When m_s tends to infinity, the observational model for $Y^{(s)}$ effectively contains close to the amount of information as the observational model in which $X^{(s)}$ is directly observed. Following this intuition, the right information quantity to consider for servers with large m_s is the Fisher information of the observational model for $X^{(s)} = (X_i^{(s)})_{i \in [n_s]}$. Through data processing arguments, we can pass to the corresponding conditional Fisher information after privatization. We consider the following dynamics for $X^{(s)}$:

$$X_i^{(s)} = f + R 2^{-L(\alpha+1/2)-1} \sum_{k=1}^{2^L} (1/2 - B_{ki}^{(s)}) \phi_{Lk}, \quad i \in [n_s], \tag{19}$$

for i.i.d. Beta$(8,8)$ random variables $B_{ki}^{(s)}$. By symmetry, the centered variables $U_{ki}^{(s)} := 1/2 - B_{ki}^{(s)}$ have mean zero and satisfy $|U_{ki}^{(s)}| \leq 1/2$, so each $X_i^{(s)}$ takes values in a Hölder ball of radius R and has mean f. Writing $S_f^{X^{(s)}}$ for the score of this latent $X^{(s)}$-experiment in the finite-dimensional submodel $\mathcal{H}_{R,L}^{\alpha}$, we let $I_f^{X^{(s)}|T^{(s)}}$ denote the corresponding conditional Fisher information after the transcript,

$$I_f^{X^{(s)}|T^{(s)}} = \mathbb{E}_f \, \mathbb{E}_f \left[S_f^{X^{(s)}} \middle| T^{(s)} \right] \mathbb{E}_f \left[S_f^{X^{(s)}} \middle| T^{(s)} \right]^\top. \tag{20}$$

The corresponding non-private Fisher information is denoted by $I_f^{X^{(s)}} = \mathbb{E}_f S_f^{X^{(s)}} S_f^{X^{(s)}\top}$. The exact normalization of the X-score and its relation to the coefficient parameter are given in Section F.1, specifically around (22) and Lemma 11.

The crux of the proof lies in the following lemma, which provides a lower bound on the minimax risk by expressing it in terms of a combination of Fisher information quantities—each capturing a different bottleneck (sample sizes, number of design points, privacy constraints).

Lemma 3. *Let $L \in \mathbb{N}$. For each augmentation profile $\eta = (\eta_1, \ldots, \eta_S) \in \{Y, X\}^S$, define*

$$I_f^{\eta_s|T^{(s)}} := \begin{cases} I_f^{Y^{(s)}|T^{(s)}}, & \eta_s = Y, \\ I_f^{X^{(s)}|T^{(s)}}, & \eta_s = X. \end{cases}$$

Then the minimax risk $\inf_{\hat{f}\in\mathcal{M}_{\varepsilon,\delta}} \sup_{f\in\mathcal{H}^\alpha(R)} \mathbb{E}_f\|\hat{f}-f\|_2^2$ *is lower bounded by*

$$\frac{2^{2L}}{\min_{\eta\in\{Y,X\}^S} \sup_{f\in\mathcal{H}^\alpha_{R,L}} \sum_{s=1}^S \mathrm{Tr}\big(I_f^{\eta_s|T^{(s)}}\big) + 2^{2L(\alpha+1)}}. \tag{21}$$

It can be seen as a generalization of the technique introduced by [10], which allows us to account for the situation where the observational model for $Y^{(s)}$ leads to nearly the same performance for estimating f as if $X^{(s)}$ were observed directly. Its proof relies on a multivariate version of the von Trees inequality of [20] and data processing arguments exploiting the Markov chain structure $f \to X^{(s)} \to Y^{(s)}$ of the model.

After combining Lemma 3 with the uniform serverwise bounds in Lemma 4, and then minimizing over augmentation profiles server by server, the denominator reduces to the same four bottleneck quantities as before.

(a) In the first scenario we consider, both the privacy constraint, the limited number of observations as well the limited number of measurements affect the information that the transcript $T^{(s)}$ contains on f. This corresponds with the quantity $I_f^{Y^{(s)}|T^{(s)}}$, which we in turn bound by a constant multiple of $\sigma_s^{-2} m_s n_s^2 \varepsilon_s^2$.

(b) In the second scenario, the privacy constraint is relatively lenient for the s-th server, in which case the Fisher information of its transcript satisfies $I_f^{Y^{(s)}|T^{(s)}} \asymp I_f^{Y^{(s)}}$, the latter of which can be computed to equal $\sigma_s^{-2} 2^L m_s n_s$.

(c) In the scenario that the s-th server contains relatively many measurements per individual, the sample size n_s and privacy constraint form the bottleneck for the s-th server. In this case, $I_f^{X^{(s)}|T^{(s)}}$ is the relevant quantity to consider, which we bound by a constant multiple of $2^{L(2\alpha+1)} n_s^2 \varepsilon_s^2$ using data processing arguments.

(d) In the last scenario, both measurements and privacy budget are no longer part of the bottleneck, and just the (number of) X_i observations locally determine information transmitted by machine s. In this case, the $I_f^{X^{(s)}|T^{(s)}} \asymp I_f^{X^{(s)}} \le 2^{2L(\alpha+1)} n_s$.

The bounds on the traces of the various Fisher information are given in the Lemma 4 below, for which we give a proof in Section F.1 of the Supplementary Material.

Lemma 4. *Suppose that for every* $s \in [S]$, $\delta_s \log(1/\delta_s) \lesssim \left(\frac{n_s}{n_s \vee m_s} \varepsilon_s^2 D_*^{-1}\right) \wedge \left(\varepsilon_s^3 n_s^{3/4} D_*^{-3/2}\right)$. *Then, uniformly over* $f \in \mathcal{H}^\alpha_{R,L}$,

$$\mathrm{Tr}(I_f^{Y^{(s)}|T^{(s)}}) \lesssim \sigma_s^{-2} m_s n_s^2 \varepsilon_s^2, \quad \mathrm{Tr}(I_f^{Y^{(s)}}) \lesssim \sigma_s^{-2} 2^L m_s n_s,$$
$$\mathrm{Tr}(I_f^{X^{(s)}|T^{(s)}}) \lesssim 2^{L(2\alpha+1)} n_s^2 \varepsilon_s^2, \quad \mathrm{Tr}(I_f^{X^{(s)}}) \lesssim 2^{2L(\alpha+1)} n_s.$$

Combining Lemma 3 with Lemma 4, together with the basic data-processing bounds $\mathrm{Tr}(I_f^{Y^{(s)}|T^{(s)}}) \le \mathrm{Tr}(I_f^{Y^{(s)}})$ and $\mathrm{Tr}(I_f^{X^{(s)}|T^{(s)}}) \le \mathrm{Tr}(I_f^{X^{(s)}})$, yields a lower bound in terms of the

resolution level L. Choosing $2^L \asymp D_*$, where D_* solves (2), gives the claimed rate $D_*^{-2\alpha}$. In this way, the independent-design lower bound is driven by whichever of the four information bottlenecks dominates at a given server, and the overall rate emerges from aggregating these local constraints across the federation. The full proof is given in Section F.1.

This perspective also clarifies the contrast with the common-design setting considered next. There, the shared grid changes the structure of the relevant bottlenecks and leads to a simpler lower-bound argument, since the discretization constraint enters in a fundamentally different way.

4.2 Lower bound for common design

For the common design setting, we prove the following lower bound on the minimax risk.

Theorem 6. *Let D_* solve* (5)*. Assume that the common design satisfies* (3)*, that* (4) *holds, and that, for every $s \in [S]$, $\delta_s \log(1/\delta_s) \lesssim D_*^{-1}\varepsilon_s^2$. Then*

$$\inf_{\hat{f} \in \mathcal{M}_{\varepsilon,\delta}} \sup_{f \in \mathcal{H}^\alpha(R)} \mathbb{E}_f \|\hat{f} - f\|_2^2 \gtrsim D_*^{-2\alpha}, \tag{22}$$

The full proof is given in Section F.2.

Below, we briefly sketch the argument, emphasizing the main distinction from the independent-design lower bound. A detailed proof is deferred to the Supplementary Material. In the common-design setting, the lower bound splits into two regimes. When $D_* = m$, the rate is driven by the deterministic approximation error $m^{-2\alpha}$, exactly as in the classical non-private common-design problem.

In the privacy-driven regime, we use a direct finite-dimensional submodel argument based on the observed vectors $Y_i^{(s)}$. More precisely, we construct a D-dimensional submodel, with $D \asymp D_*$, using disjoint smooth bump functions placed on a quasi-uniform common design satisfying (3), and apply a multivariate van Trees inequality to this submodel. This reduces the problem to controlling, for each server s, the Fisher information in the non-private Y-model together with its privatized counterpart after the server-s differential-privacy channel, namely

$$\mathrm{Tr}(I_f(Y^{(s)})) \qquad \text{and} \qquad \mathbb{E}\mathrm{Tr}(\boldsymbol{C}_f^{(s)}(T^{(s)})),$$

where $\boldsymbol{C}_f^{(s)}(T^{(s)}) = \mathbb{E}[S_f^{Y^{(s)}} \mid T^{(s)}]\mathbb{E}[S_f^{Y^{(s)}} \mid T^{(s)}]^\top$. The first term is of order $\sigma_s^{-2} D^2 n_s$, while the second is bounded by $\sigma_s^{-2} D n_s^2 \varepsilon_s^2$ via the differential-privacy data-processing argument. Optimizing over D then yields the privacy-dependent part of the lower bound.

5 Numerical Study

5.1 Simulation study

We evaluate the proposed estimator in a *federated privacy* setting under two sampling designs for functional data: a *common design*, where all individuals are observed on the same grid, and an

independent design, where each individual is observed on an independent random grid. In both settings, each individual i has an underlying random signal f_i drawn from a Hölder/Besov-type wavelet model with smoothness $\alpha = 2$ and radius $R = 2$, and observations are contaminated with Gaussian noise with $\sigma = 1$. The inferential target is the population mean function $f(t) = \mathbb{E}[f_i(t)]$. Performance is measured by mean squared error (MSE). We summarize a practical tuning rule here, while fuller implementation details are deferred to Appendix B.

Code for the simulation experiments and empirical illustration is available at `https://github.com/abhinavc3/Federated-Functional-Mean-Estimation`.

In practice, the tuning is fairly simple. For the common-design estimator, the main inputs are the working smoothness level α_{tune} and radius R_{tune}, which determine the grouping and smoothing scales. Our sensitivity analysis suggests that this estimator is relatively stable over a broad range of α_{tune}, provided R_{tune} is not chosen excessively large. For the independent-design estimator, one additionally chooses the clipping constant $C_{\tau,\text{tune}}$, and the procedure is somewhat more sensitive to under-specifying smoothness. A practical rule is therefore to choose a moderately conservative smoothness level, erring slightly on the high side for α_{tune}, and to avoid overly large clipping constants. Additional implementation choices and sensitivity plots are given in Appendix B and Section B.5.

Federated privacy and replication. We consider privacy budgets $\varepsilon \in \{0.5, 2, 8\}$ with $\delta = 1/n^2$, where n denotes the per-server sample size in the given configuration. Each configuration is repeated over 50 Monte Carlo replicates. The results can be seen in Figure 1. Throughout, we compare the common-design (solid) and independent-design (dashed) implementations under the same privacy budgets.

MSE versus per-server sample size n (left panel in Figure 1). *Sweep setting.* We fix the number of servers $S = 1$ and the number of measurements per individual $m = 64$, and vary $n \in \{100, 200, 300, 450, 600\}$. *Takeaways.* As n increases, MSE decreases across all privacy levels, and the curves are ordered by privacy strength: smaller ε yields larger error. Across this range, the independent-design estimator is uniformly competitive and often substantially more accurate than the common-design estimator, with the gap most pronounced under stronger privacy ($\varepsilon = 0.5$) this is in line with our findings of Section 2.3.

MSE versus number of measurements m (middle panel in Figure 1). *Sweep setting.* We fix $S = 1$ and $n = 200$, and vary $m \in \{10, 25, 60, 100\}$. *Takeaways.* For the independent design, increasing m yields a clear reduction in MSE, consistent with improved accuracy of the underlying projection/integration step. For the common design in moderate-to-strong privacy the error plateaus relatively quickly as m increases, indicating a regime in which additional within-individual measurements no longer improve accuracy because the privacy constraint becomes the dominant bottleneck. This qualitative behavior is consistent with Corollary 2, which

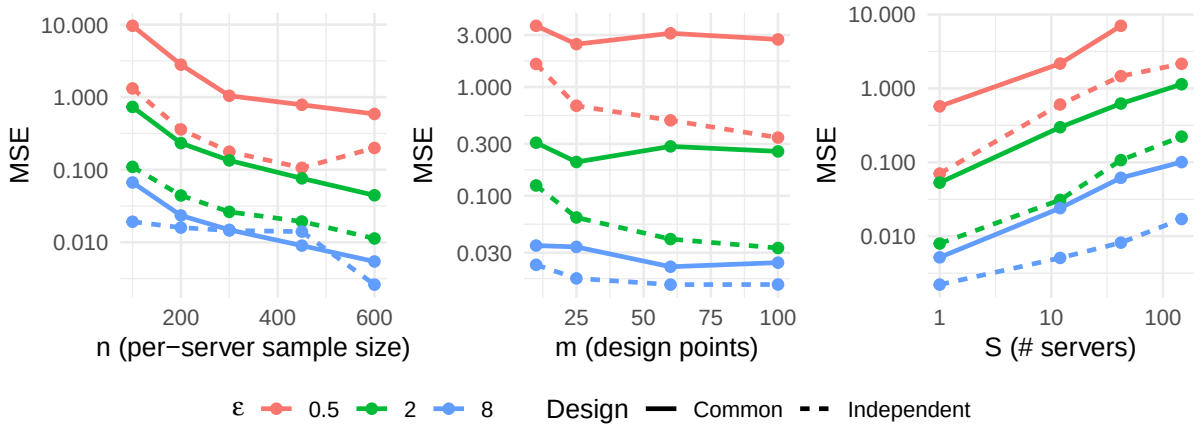


Figure 1: Mean squared error (MSE; log scale) for the common-design (solid) and independent-design (dashed) estimators under three privacy budgets (colors). Left: MSE versus per-server sample size n (with $S = 1$ and $m = 64$). Middle: MSE versus number of design points m (with $S = 1$ and $n = 200$). Right: MSE versus number of servers S at fixed total sample size $Sn = 512$ (with $m = 256$).

predicts that gains from increasing m are limited once privacy-induced noise dominates.

MSE versus number of servers S (right panel in Figure 1). *Sweep setting.* We fix an approximate total sample size $N_{\text{tot}} = 512$ and set $m = 256$. We vary the number of servers $S \in \{1, 12, 42, 147\}$ and set $n = \text{round}(N_{\text{tot}}/S)$ so that $Sn \approx N_{\text{tot}}$. *Takeaways.* Holding $Sn \approx N_{\text{tot}}$ fixed, increasing the number of servers leads to a pronounced degradation in accuracy for both designs. This illustrates the practical cost of federation under privacy: splitting a fixed number of data points across many servers reduces the per-server sample size n, which in turn increases the magnitude of the privacy noise injected at each server and worsens downstream estimation. The effect is strongest for small ε but remains visible even at moderate privacy budgets.

5.2 Real data example: federated estimation of mean BMI trajectories in HRS

We illustrate the proposed *federated-private* independent-design procedure on longitudinal body-mass index (BMI) measurements from the Health and Retirement Study (HRS), a biennial panel survey of older Americans. We treat age as the design variable and BMI as the response, and partition respondents by U.S. Census region so that each region acts as a separate server. This yields an emulated multi-server federated setting based on regional partitions, with irregular observation times, heterogeneous local sample sizes, and a natural population-level target: the mean BMI trajectory as a function of age. It is also a setting where privacy is intrinsically important, since both BMI and age-at-measurement are sensitive individual-level attributes that

respondents may reasonably wish to protect. The implementation details, including preprocessing and the estimator settings used in this analysis, are given in Appendix B.2.

Figure 2 summarizes the resulting privacy-utility tradeoff. The left panel shows the estimated mean BMI trajectory for several privacy budgets ε, together with the pooled non-private benchmark ($\varepsilon = \infty$). The right panel reports the mean squared deviation of each private estimate from that pooled non-private curve. A key qualitative feature is preserved even under privacy constraints: the estimated mean BMI trajectory retains its non-monotone shape, increasing over earlier ages in the observed range and declining at older ages. Thus, even in the private setting, the estimator recovers the main trend of the BMI-age relationship rather than only a heavily distorted version of it.

At the same time, the figure makes the privacy-utility tradeoff transparent. The strongest privacy regime produces visible distortion relative to the pooled non-private benchmark, whereas for moderate and large privacy budgets the private curves track the non-private trajectory much more closely. This pattern is also reflected quantitatively in the right panel, where the discrepancy from the pooled non-private curve decreases as ε increases.

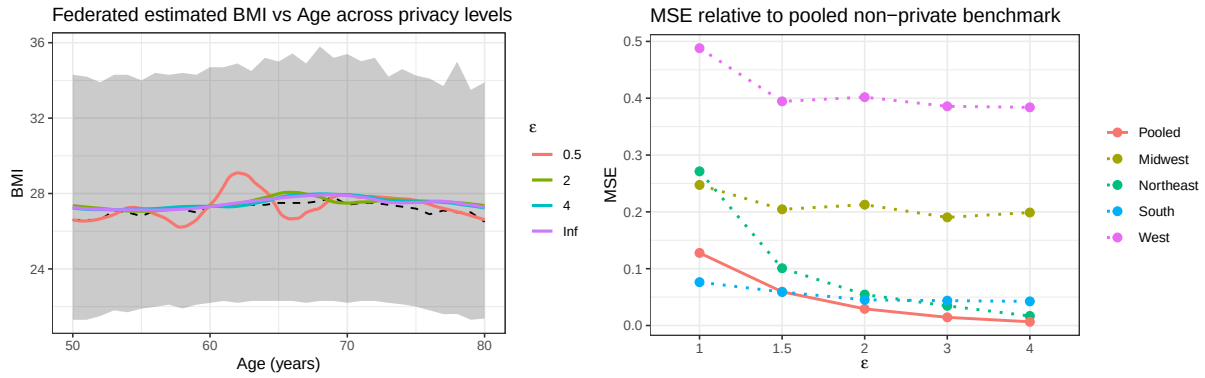
The figure also includes region-specific private curves, plotted as dotted lines, alongside the pooled federated estimator. Taken together, the HRS example shows that one can still recover a stable and interpretable population-level functional signal without pooling raw data across servers.

Overall, this real-data analysis demonstrates that the proposed method is practically viable beyond simulations. Even in an emulated multi-server longitudinal setting with irregular designs and heterogeneous servers, the federated-private estimator preserves the main qualitative trend of the mean BMI trajectory and remains reasonably close to the non-private benchmark except under the strongest privacy constraint. This supports the use of federated private functional mean estimation for population-level inference in settings where direct data pooling is not feasible.

6 Conclusion

In this paper, we studied federated functional mean estimation under heterogeneous differential privacy constraints. We established minimax upper and lower bounds for both the common-design and independent-design settings, showing that the cost of privacy depends in a fundamental way on the sampling design and the heterogeneity across servers. We also proposed one-shot private estimators that attain the optimal rates up to polylogarithmic factors and accommodate heterogeneity in sample sizes, numbers of measurements, privacy budgets, and noise levels.

Our theoretical findings are supported by an expanded simulation study and by a real-data illustration in an emulated multi-server federated setting based on the Health and Retirement Study. Together, these results show that meaningful and statistically efficient functional estimation is possible under federated privacy constraints, while also highlighting clear differences



(a) (b)

Figure 2: HRS application. Left: estimated mean BMI trajectory as a function of age under the federated-private independent-design estimator for several privacy budgets ε, with $\varepsilon = \infty$ as the pooled non-private benchmark. Right: mean squared deviation from the pooled non-private curve versus ε. The solid line is the pooled federated estimator, and the dotted lines are region-specific estimators fit separately within each Census region; each point averages over $B = 50$ reruns of the privatization mechanism.

between common and independent design. Several interesting directions remain for future work, including adaptive procedures for unknown smoothness (using ideas from [11]) and extensions to broader privacy-constrained functional problems such as functional regression.

Data Availability

Code and data for the simulation experiments and empirical illustration are available at `https://github.com/abhinavc3/Federated-Functional-Mean-Estimation`.

References

[1] Jayadev Acharya, Clément L Canonne, Aditya Vikram Singh, and Himanshu Tyagi. Optimal rates for nonparametric density estimation under communication constraints. *IEEE Transactions on Information Theory*, 2023.

[2] Jayadev Acharya, Clément L Canonne, Ziteng Sun, and Himanshu Tyagi. Unified lower bounds for interactive high-dimensional estimation under information constraints. *Advances in Neural Information Processing Systems*, 36, 2024.

[3] Jayadev Acharya, Yuhan Liu, and Ziteng Sun. Discrete distribution estimation under user-level local differential privacy. In *Proceedings of The 26th International Conference*

on Artificial Intelligence and Statistics, volume 206, pages 8561–8585. PMLR, 25–27 Apr 2023.

[4] Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory*, pages 48–78. PMLR, 2021.

[5] Chiara Amorino and Arnaud Gloter. Minimax rate for multivariate data under componentwise local differential privacy constraints. *The Annals of Statistics*, 53(3):1176–1202, 2025.

[6] Arnab Auddy, T Tony Cai, and Abhinav Chakraborty. Minimax and adaptive transfer learning for nonparametric classification under distributed differential privacy constraints. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkaf070, 2025.

[7] Leighton Pate Barnes, Wei-Ning Chen, and Ayfer Özgür. Fisher information under local differential privacy. *IEEE Journal on Selected Areas in Information Theory*, 1(3):645–659, 2020.

[8] Francoise Beaufays, Kanishka Rao, Rajiv Mathews, and Swaroop Ramaswamy. Federated learning for emoji prediction in a mobile keyboard. *arXiv:1906.04329*, 2019.

[9] T. Tony Cai, Abhinav Chakraborty, and Lasse Vuursteen. Federated nonparametric hypothesis testing with differential privacy constraints: Optimal rates and adaptive tests. 2024.

[10] T Tony Cai, Abhinav Chakraborty, and Lasse Vuursteen. Optimal federated learning for nonparametric regression with heterogeneous distributed differential privacy constraints. *arXiv preprint arXiv:2406.06755*, 2024.

[11] T Tony Cai, Abhinav Chakraborty, and Lasse Vuursteen. The cost of adaptation under differential privacy: Optimal adaptive federated density estimation. *arXiv preprint arXiv:2512.14337*, 2025.

[12] T. Tony Cai and Ming Yuan. Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics*, 39(5):2330 – 2355, 2011.

[13] Clément L Canonne and Yucheng Sun. Private distribution testing with heterogeneous constraints: Your epsilon might not be mine. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, 2024.

[14] Abhinav Chakraborty, Arnab Auddy, and T Tony Cai. When data can't meet: Estimating correlation across privacy barriers. *Advances in Neural Information Processing Systems*, 38:95086–95120, 2026.

[15] Yuval Dagan, Michael I. Jordan, Xuelin Yang, Lydia Zakynthinou, and Nikita Zhivotovskiy. Dimension-free private mean estimation for anisotropic distributions. 2024.

[16] Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

[17] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.

[18] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[19] Frédéric Ferraty and Philippe Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Science & Business Media, Berlin, Heidelberg, 2006.

[20] Richard D Gill and Boris Y Levit. Applications of the van trees inequality: a bayesian cramér-rao bound. *Bernoulli*, pages 59–79, 1995.

[21] Rob Hall, Alessandro Rinaldo, and Larry Wasserman. Differential privacy for functions and functional data. *The Journal of Machine Learning Research*, 14(1):703–727, 2013.

[22] Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129. Springer Science & Business Media, 2012.

[23] Akira Horiguchi, Li Ma, and Botond T Szabó. Sampling depth trade-off in function estimation under a two-level design. *arXiv preprint arXiv:2310.02968*, 2023.

[24] Clément Lalanne, Aurélien Garivier, and Rémi Gribonval. On the statistical complexity of estimation and testing under privacy constraints. *Transactions on Machine Learning Research Journal*, 2023.

[25] Daniel Levy, Ziteng Sun, Kareem Amin, Satyen Kale, Alex Kulesza, Mehryar Mohri, and Ananda Theertha Suresh. Learning with user-level privacy. *Advances in Neural Information Processing Systems*, 34:12466–12479, 2021.

[26] Yuhan Liu, Ananda Theertha Suresh, Felix Xinnan X Yu, Sanjiv Kumar, and Michael Riley. Learning discrete distributions: user vs item-level privacy. *Advances in Neural Information Processing Systems*, 33:20965–20976, 2020.

[27] Ardalan Mirshani, Matthew Reimherr, and Aleksandra Slavković. Formal privacy for functional data with gaussian perturbations. In *International Conference on Machine Learning*, pages 4595–4604. PMLR, 2019.

[28] Shyam Narayanan, Vahab Mirrokni, and Hossein Esfandiari. Tight and robust private mean estimation with few users. In *International Conference on Machine Learning*, pages 16383–16412. PMLR, 2022.

[29] JO Ramsay and BW Silverman. Principal components analysis for functional data. *Functional data analysis*, pages 147–172, 2005.

[30] John A Rice and Bernard W Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):233–243, 1991.

[31] Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer series in statistics. Springer, New York ; London, 2009. OCLC: ocn300399286.

[32] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and its application*, 3(1):257–295, 2016.

[33] Gengyu Xue, Zhenhua Lin, and Yi Yu. Optimal estimation in private distributed functional data analysis. 2024.