Optimal Federated Learning for Nonparametric Regression with Heterogeneous Distributed Differential Privacy Constraints

T. Tony Cai* Abhinav Chakraborty* Lasse Vuursteen*

tcai@wharton.upenn.edu, abch@wharton.upenn.edu, lassev@wharton.upenn.edu

Abstract

This paper studies federated learning for nonparametric regression with distributed samples across servers, each adhering to distinct differential privacy constraints. The setting is heterogeneous, with varying sample sizes and privacy constraints across servers. We consider both global and pointwise estimation, establishing optimal rates of convergence over Besov spaces.

We propose distributed privacy-preserving estimators and investigate their minimax risk, establishing minimax lower bounds (up to a logarithmic factor) for both global and pointwise estimation. Our findings are illustrated through simulations and real data examples from the National Health and Nutrition Examination Survey (NHANES), highlighting the effects of privacy budgets, number of servers, and sample sizes. The real data applications demonstrate the estimator's utility in analyzing non-linear relationships like lung function with age, and folate, vitamin B12, with homocysteine levels.

Our findings highlight the trade-off between statistical accuracy and privacy, characterizing the compromise in terms of privacy budgets and the loss from distributing data within the privacy framework as a whole. This insight captures the folklore wisdom that it is easier to retain privacy in larger samples, and explores the differences between pointwise and global estimation under distributed privacy constraints. Analogous optimality results for nonparametric density estimation are also established.

Keywords: Besov Spaces, Distributed Computation, Differential Privacy, Minimax Risk, Nonparametric Regression, Function Estimation.

1 Introduction

In today's data-driven world, the proliferation of personal data and technological advancements has made the protection of privacy a matter of paramount importance. Developing

*Department of Statistics and Data Science, University of Pennsylvania

statistical methods with privacy guarantees is becoming increasingly important. Differential privacy (DP), one of the most widely adopted privacy frameworks, ensures that statistical analysis results do not divulge any sensitive information about the input data. DP was introduced in the seminal work by Dwork et al. [29]. Since its inception, DP has garnered significant academic attention [8, 30] and notable applications within industry leaders, including Google [33], Microsoft [25], and Apple [55]. It has also been embraced by governmental entities like the US Census Bureau [1].

A common setting in many real-life applications is the distributed nature of data collection and analysis. For example, medical data is spread across various hospitals in healthcare, customer data is stored in different branches or databases in financial institutions and various modern technologies such as self driving cars rely on federated learning from networks of users, see for example [12] and Section A in the Supplementary Material for a list of references. DP has found applications in many of these domains relating to, for example, healthcare, finance, tech and social sciences, where preserving individuals' data privacy is of utmost concern. In such scenarios, it is vital to develop efficient estimation techniques that respect privacy constraints while harnessing the collective potential of distributed data.

Federated learning is a machine learning paradigm designed to address the challenges of data governance and privacy. It enables organizations or groups, whether from diverse geographic regions or within the same organization, to collaboratively train and improve a shared global statistical model without external sharing of raw data. The learning process occurs locally at each participating entity, which we shall refer to as *servers*. The servers exchange only characteristics of their data, such as parameter estimates or gradients, in a way that preserves privacy of the individuals comprising their data. Federated learning facilitates secure collaboration across industries like retail, manufacturing, healthcare, and financial services, allowing them to harness the power of data analysis while upholding data privacy and security.

Rigorous study of theoretical performance in federated learning settings with infor-

mation and communication constraints has been conducted in, for example, bandwidth constraint problems of which we provide a brief overview in Section A of the Supplementary Material. Under DP constraints, theoretical performance in federated learning settings have been studied for various parametric estimation and testing problems [48, 5, 47, 51]. Federated learning settings where each server's sample consists of one individual observation (referred to as *local* differential privacy settings) have been studied in many-normal-means model, discrete distributions and parametric models [27, 28, 10, 2, 57], nonparametric density estimation [53, 44, 16] and non-parametric regression setting [14, 35].

This paper investigates the statistical optimality of federated learning under a novel privacy framework, Federated Differential Privacy (FDP), in the context of nonparametric regression. We consider a setting where data is distributed among entities, such as hospitals, that are concerned about sharing data due to privacy concerns. Each entity communicates a transcript adhering to distinct DP requirements under FDP, and we assume a scenario with m servers, each with n_j observations, where $j = 1, \ldots, m$. Our framework provides an intermediate privacy model between central and local DP.

Our goals are two-fold: first, we establish optimal rates of convergence, measured in terms of minimax risk, for estimating the nonparametric regression function under FDP constraints; second, we construct a rate-optimal estimator in this setting. We also propose general information-theoretic lower bound techniques that demonstrate the optimality of our results. These techniques are broadly applicable and could be useful in establishing lower bounds for other problems beyond nonparametric regression, which may be of independent interest to readers. We explore both global and pointwise estimation, providing quantifiable measures of the trade-off between accuracy and privacy. Recognizing that global estimation behaves differently than pointwise estimation under classical settings [19], we characterize how FDP constraints impact global and pointwise estimation risks.

To further validate our theoretical findings, we conducted extensive simulation studies and real-world experiments using data from the National Health and Nutrition Exami-

nation Survey (NHANES). These empirical studies demonstrate the performance of our estimators within the FDP framework, comparing their accuracy and privacy trade-offs against classical non-private methods. Specifically, the real data applications examine important nonlinear relationships, such as those between lung function and age, and between folate, vitamin B12, and homocysteine levels. These examples illustrate how privacy mechanisms impact estimation accuracy in practical settings, demonstrating how our approach effectively balances privacy preservation with statistical accuracy in real-world applications.

1.1 Problem formulation

We will begin by formally introducing the general framework of distributed estimation under privacy constraints. Consider a family of probability measures $\{P_f\}_{f\in\mathcal{F}}$ on the measurable space $(\mathcal{Z}, \mathscr{Z})$, parameterized by $f \in \mathcal{F}$. We consider a setting where $N = \sum_{j=1}^{m} n_j$ i.i.d. observations are drawn from a distribution P_f and distributed across m servers. Each server $j = 1, \ldots, m$ holds n_j observations.

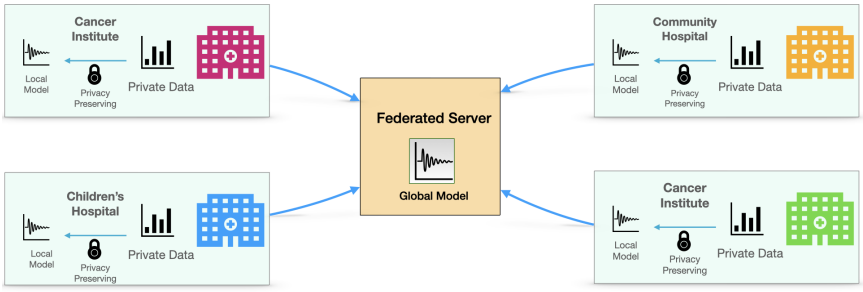


Figure 1: An illustration of the federated learning framework.

Let us denote by $Z^{(j)} = \{Z_i^{(j)}\}_{i=1}^{n_j}$ the n_j realizations from P_f on the j-th server. For each server, we output a (randomized) transcript $T^{(j)}$ based on $Z^{(j)}$, where the law of the transcript is given by a distribution conditionally on $Z^{(j)}$, $\mathbb{P}(\cdot|Z^{(j)})$ on a measurable space $(\mathcal{T}, \mathscr{T})$. The transcript $T = (T^{(j)})_{j=1}^m$ has to satisfy $(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) \equiv (\varepsilon_j, \delta_j)_{j=1}^m$-federated differentially privacy (FDP) constraint, which is defined as follows.

Definition 1.1. The transcript $T = (T^{(j)})_{j=1}^m$ is $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$-federated differentially private (FDP)

if for all $j \in [m]$, $A \in \mathscr{T}$[1] and $z, z' \in \mathcal{Z}^{n_j}$ differing in one individual datum it holds that

$$\mathbb{P}\left(T^{(j)} \in A | Z^{(j)} = z\right) \leq e^{\varepsilon_j} \mathbb{P}\left(T^{(j)} \in A | Z^{(j)} = z'\right) + \delta_j.$$

In the above definition,"differing in one datum" refers to being Hamming distance "neighbors." Specifically, local datasets $Z^{(j)}$ and $\tilde{Z}^{(j)}$ are *neighboring* if their Hamming distance is at most 1, calculated over $\mathcal{Z}^{n_j} \times \mathcal{Z}^{n_j}$. In other words, $\tilde{Z}^{(j)}$ can be derived from $Z^{(j)}$ by modifying at most one observation among $Z^{(j)}1, \ldots, Z^{(j)}n_j$. The smaller the values of ε_j and δ_j, the stricter the privacy constraint. We consider $\varepsilon_j \leq C_\varepsilon$ for $j = 1, \ldots, m$, with a fixed constant $C_\varepsilon > 0$ that does not affect the derived rates.

The FDP framework applies to situations where sensitive data is held by multiple parties, each generating an output while ensuring differential privacy. Within such a distributed protocol, the transcripts from each server depend only on its local data, with no information exchanged between servers. This occurs, for example, when multiple trials concerning the same population are conducted, but each location (e.g. hospital) does not wish to pool their original data due to privacy concerns. The framework encapsulates the commonly studied local DP setting ($n_j = 1$), where privacy mechanisms are applied at the individual level, as well as the central DP setting ($m = 1$), as special cases.

Each server transmits its transcript to the central server. The central server, utilizing all transcripts $T := (T^{(1)}, \ldots, T^{(m)})$, computes an estimator $\hat{f} : \mathcal{T}^m \to \mathcal{F}$. We refer to the pair $(\hat{f}, \{(\mathbb{P}(\cdot|z))_{z \in \mathcal{Z}}\}_{j=1}^m)$ as a *distributed estimation protocol*, which we shall sometimes just denote as $\hat{f}$. We denote the vector of the differing DP levels by $(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) = \{(\varepsilon_j, \delta_j)\}_{j=1}^m$ and denote the class of *distributed estimation protocols*, i.e. $\left(\hat{f}, \{(\mathbb{P}(\cdot|z))_{z \in \mathcal{Z}}\}_{j=1}^m\right)$ satisfying Definition 1.1, with $\mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$. We let $\mathbb{P}_f$ denote the joint law of transcripts and the $N = \sum_{j=1}^m n_j$ i.i.d. observations generated from P_f. We let $\mathbb{E}_f$ denote the expectation corresponding to $\mathbb{P}_f$.

In the context of nonparametric regression, the distributed estimation problem arises

[1]Our lower bound results hold for transcripts in standard Borel spaces. Larger sigma-algebras can be considered (making privacy constraints more strict), as long as the quantities in proofs remain measurable.

when data is distributed among multiple servers. Specifically, for each server j, the data $Z^{(j)} = \{(Y_i^{(j)}, X_i^{(j)})\}_{i=1}^{n_j}$ consists of n_j pairs of observations $(Y_i^{(j)}, X_i^{(j)})$. Here, $X_i^{(j)}$ represents the input variable, and $Y_i^{(j)}$ represents the corresponding response variable. We assume that under P_f, $X_i^{(j)}$ and $Y_i^{(j)}$ are generated by the relationship

$$Y_i^{(j)} = f\left(X_i^{(j)}\right) + \xi_i^{(j)}, \quad X_i^{(j)} \sim U[0,1]. \tag{1}$$

Here, f is an unknown function representing the underlying relationship between the input and response variables. The term $\xi_i^{(j)}$ represents random noise, assumed to be independent of $X_i^{(j)}$, and follows a Gaussian distribution with mean zero and known variance, which we assume equal to one without loss of generality. The assumption that $X_i^{(j)} \sim U[0,1]$ is made for simplicity and can be relaxed, see Section D in the Supplementary Material for a discussion on the general case.

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, for global estimation, 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 (under P_f) and construction of the transcripts. As in the conventional framework, a quantity of particular interest in federated learning is the *global minimax risk* for the distributed private protocols over function class $\mathcal{F}$,

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

The global risk characterizes the difficulty of the distributed learning problem over the function class $\mathcal{F}$ when trying to infer the entire function underlying the data whilst adhering to the heterogeneous privacy constraints.

Besides global estimation, it is also of interest to estimate f at a fixed point $x_0 \in (0,1)$ under the mean squared error (MSE). The *pointwise minimax risk* in that case is given by

$$\inf_{\hat{f} \in \mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})} \sup_{f \in \mathcal{F}} \mathbb{E}_f (\hat{f}(x_0) - f(x_0))^2, \quad \text{for } x_0 \in (0,1), \tag{3}$$

where $\hat{f}(x_0)$ denotes the estimated function value at $x_0 \in (0,1)$. The pointwise risk is particularly useful in understanding the behavior of estimators at specific points within the domain, which can be crucial in applications where certain regions are of particular interest or have higher consequences associated with estimation errors. It is known that in the classical setting, without privacy constraints, there are important differences between the global risk and pointwise risk in terms of performance. See, for example, [18].

We consider estimating f over the Besov ball of radius $R > 0$, denoted as $\mathcal{B}_{p,q}^{\alpha,R}[0,1]$ (defined in (S.9) in the supplement), where $p \geq 2$, $q \geq 1$ and $\alpha - 1/p > 1/2$. This Besov space offers a suitable framework for analyzing functions with specific smoothness characteristics. Operating within this space allows us to encompass diverse function classes, accommodating varying levels of smoothness and complexity.

1.2 Main contribution

We highlight our main contributions as follows:

- **Optimal Nonparametric Estimation under FDP:** We study the cost of differential privacy in federated learning for nonparametric regression, introducing DP estimators for the global and pointwise risks. We derive theoretical performance guarantees with matching minimax lower bounds, establishing their optimality. These results reveal the principle phenomena in federated settings with varying privacy budgets and sample sizes, capturing the cost of distributing data.

- **General Information-Theoretic Optimality Techniques:** We develop general information-theoretic techniques to establish optimality results within the FDP frame-

work. Whilst the majority of existing literature studying DP is divided into central DP and local DP, our framework fully generalizes the unit of privacy and allows for a unified treatment of these settings. Consequently, our FDP results also establish the minimax rates for central DP and local DP constraints nonparametric regression (up to log-factors) and tighten existing minimax rates for density estimation, as special cases.

- **Empirical Studies:** To further support our theoretical findings, we conduct extensive simulation studies and illustrate our methods on real-world data. These empirical studies demonstrate the performance of our estimators within the FDP framework, comparing their accuracy and privacy trade-offs against classical DP methods. Specifically, we apply our approach to data from the National Health and Nutrition Examination Survey (NHANES), analyzing relationships such as lung function and age, as well as folate, vitamin B12, and homocysteine levels.

Next, we briefly elaborate on the key theoretical results. We quantify the cost of differential privacy in the federated setting for both the minimax global risk given by (2) and the pointwise risk as in (3). To achieve this, we introduce two federated differentially private estimators – one for global and one for pointwise estimation. We obtain matching minimax lower bounds, up to logarithmic factors, thereby establishing their optimality.

Our analysis uncovers intriguing phenomena that go unnoticed in settings where servers are assumed to have homogeneous privacy budgets. Further discussion on these broader findings is deferred to Section 3.1. The results for the homogeneous case, where privacy budgets are equal among servers ($\varepsilon_j = \varepsilon$, $\delta_j = \delta$, and $n_j = n$ for $j = 1, \ldots, m$), yield novel insights. In this case, our results yield the following minimax rate for global estimation,

$$\inf_{\hat{f} \in \mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})} \sup_{f \in \mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f \|\hat{f} - f\|_2^2 \asymp \min\left\{ M_{m,n} \cdot \left((mn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+2}} + (mn)^{-\frac{2\alpha}{2\alpha+1}} \right), 1 \right\}, \qquad (4)$$

where $M_{m,n} \geq 1$ is a sequence at most of the order $\log(mn) \cdot \log(1/\delta)$. The rate $(mn)^{-\frac{2\alpha}{2\alpha+1}}$

is the minimax rate for the global risk in the unconstrained problem, and is attained whenever $n\varepsilon^2 \gtrsim (mn)^{\frac{1}{2\alpha+1}}$. The unconstrained optimal rate is attainable (up to a possibly poly-logarithmic factor) under DP constraints in the homogeneous setting as long as $n\varepsilon^2 \gtrsim (mn)^{\frac{1}{2\alpha+1}}$. Whenever $n\varepsilon^2 \ll (mn)^{\frac{1}{2\alpha+1}}$, the first term dominates and the minimax rate becomes $(mn^2\varepsilon^2)^{-\frac{2\alpha}{2\alpha+2}}$. As expected in this regime, a smaller ε, which indicates a stronger privacy guarantee, results in an larger minimax estimation error. Whenever $\varepsilon \ll (\sqrt{mn})^{-1}$, consistent estimation ceases to be possible altogether.

When $n > 1$, the different powers with which n and m appear in the minimax rate reveal an important difference between the general distributed setting and local DP; if one distributes $N = mn$ observations across m machines, the task becomes more challenging as the N observations are spread over a greater number of machines, rather than having a large number of observations on a smaller number of machines. This phenomenon has an intuitive explanation; it is easier to retain privacy in larger samples, as each individual's data will have only a small influence on the aggregate statistics of interest.

For pointwise estimation, we establish the minimax rate in the homogeneous setting;

$$\inf_{\hat{f}\in\mathcal{M}(\boldsymbol{\varepsilon},\boldsymbol{\delta})} \sup_{f\in\mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f|\hat{f}(x_0) - f(x_0)|^2 \asymp \min\left\{M_{m,n}\cdot\left(\left(mn^2\varepsilon^2\right)^{-\frac{2\nu}{2\nu+2}} + (mn)^{-\frac{2\nu}{2\nu+1}}\right), 1\right\}, \quad (5)$$

where $M_{m,n} \geq 1$ is a sequence at most of the order $\log(mn)$. The rate reveals similar phenomena as the one for the global risk above, where for $n = 1$ we recover the known minimax rate for the problem of nonparametric density estimation for the pointwise risk under local DP constraints as studied in [44]. An important difference is the quantity $\nu = \alpha - 1/p$ appearing in the exponent instead of α. This implies that privacy constraints impact pointwise estimation differently than global estimation, with the Besov parameter p influencing both the relative privacy cost and the distribution of the $N = mn$ observations, as discussed further in Section 3.1.

The estimation and lower bound techniques developed in our paper enables us to address the related problem of nonparametric density estimation also, allowing us to derive the

minimax rates displayed in (4) and (5) for global and pointwise minimax rates for the setting of FDP density estimation as well. The minimax rates derived by our technique for the global and pointwise risks in the density estimation setting match exactly (see Theorem D.1). Due to space limitations, this discussion is deferred to Section D of the Supplementary Material. Notably, our results improve over the existing central DP ($m = 1$) results of [45] and local DP ($n = 1$) results of [16] and [53], which exhibit a logarithmic gaps between their upper and lower bounds.

Our findings provide key insights for developing federated learning algorithms that balance FDP with accuracy, optimizing privacy-accuracy trade-offs. This study advances understanding of privacy-preserving machine learning in distributed settings.

1.3 Related Work

The nonparametric regression setting considered in this work bears relationships with that of nonparametric density estimation as studied in the privacy setting for global risk [28, 53, 16] and pointwise risk [44]. The aforementioned papers consider the setting of local DP, in which the privacy protection is applied at the level of individual data entries or observations. This corresponds to the case wherein $n_j = 1$ for $j = 1, \ldots, m$ in our setting.

Federated DP as considered in this paper, where DP applies at the level of the local sample consisting of multiple observations, has been studied in the homogeneous setting for discrete distributions [48, 5] and parametric mean estimation [47, 51]. [23] considers discrete distribution testing in a two server setting ($m = 2$) with differing DP constraints.

Settings in which the full data is assumed to be on a single server (i.e. $m = 1$), where a single privacy constraint applies to all the observations, have also been studied for various parametric high-dimensional problems [54, 32, 11, 41, 42, 21, 50]. The problem of mean estimation with a single server having heterogeneous privacy constraints for each individual observation have been studied in [34, 24].

Regarding existing lower bound techniques, several approaches have been developed

specifically for private settings. For instance, [46] and [7] explore private versions of general techniques, such as Fano, Assouad, and Le Cam methods, for establishing lower bounds in the central DP setting. In contrast, [10] develops Van Trees-based lower bounds for local DP, which do not extend directly to central DP. Similarly, the techniques in [3, 4] are tailored to the local DP setting, with no straightforward extension to central DP.

1.4 Organization of the paper

The rest of the paper is organized as follows. We conclude this section with notation, definitions, and assumptions. In Section 2, we present distributed estimation procedures achieving optimal global and pointwise risks under distributed privacy constraints, along with an upper bound on their statistical performance. The matching minimax lower bounds for global and pointwise risks are established in Section 3. Section 4 presents simulation studies and real-world data experiments to validate our theoretical findings.

In the Supplementary Material [20] to the article, we provide a discussion of future directions (Section B), additional results concerning the heterogeneous setting (Section C) and density estimation (Section D), additional simulations and real data examples (Section E), as well as detailed proofs of our main results.

1.5 Notation, definitions and assumptions

Throughout this article, let $N := \sum_{j=1}^{m} n_j$ and consider asymptotics in m, n_j, and the privacy budget $(\boldsymbol{\epsilon}, \boldsymbol{\delta}) := \{\varepsilon_j, \delta_j\}_{j=1}^{m}$, assuming $N \to \infty$. For positive sequences a_k, b_k, we write $a_k \lesssim b_k$ if $a_k \leq C b_k$ for some universal constant C, and $a_k \asymp b_k$ if $a_k \lesssim b_k$ and $b_k \lesssim a_k$. We denote $a_k \ll b_k$ when $a_k / b_k = o(1)$.

We use $a \vee b$ and $a \wedge b$ for the maximum and minimum of a and b, respectively. For $k \in \mathbb{N}$, $[k]$ denotes the set $\{1, \ldots, k\}$. Throughout, c and C are universal constants that may change from line to line. The Euclidean norm of $v \in \mathbb{R}^d$ is $\|v\|_2$, and for $M \in \mathbb{R}^{d \times d}$, $\|M\|$ is the spectral norm and $\mathrm{Tr}(M)$ its trace. Let I_d be the $d \times d$ identity matrix.

We assume $\nu := \alpha - 1/p > 1/2$, a necessary condition for Besov space estimation (see [40]), and let $\mathcal{B}_{p,q}^{\alpha,R}$ denote the Besov ball of radius R: $\{f \in \mathcal{B}_{p,q}^{\alpha}[0,1] : \|f\|_{\mathcal{B}_{p,q}^{\alpha}} \leq R\}$, where $R > 0$ is constant (see Section F in the Supplementary Material for details).

2 Optimal Distributed Private Estimators

This section presents the construction of optimal distributed estimators under differential privacy constraints. The estimator uses wavelet approximations up to a limited resolution level. Section 2.1 briefly introduces wavelets. In Section 2.2, we construct the estimator and provide theoretical guarantees for the global risk. Section 2.3 adapts the procedure for optimal pointwise risk.

Wavelets are known to have many favourable properties when using them for function estimation in classical settings, see for example [26, 38, 17]. Under DP constraints, wavelet constructions have other desirable properties: they allow for exact control of the estimator's *sensitivity* to changes in the data. Loosely speaking, this allows us to control the "influence" each individual observation has on the outcome of the estimator, whilst retaining the information the full sample has to a large extent.

2.1 Wavelets and Besov spaces

In nonparametric regression, we aim to construct an optimal estimator for an unknown function f based on distributed data, assuming f belongs to the Besov space $\mathcal{B}_{p,q}^{\alpha}$. Roughly, $\mathcal{B}_{p,q}^{\alpha}$ contains functions with α bounded derivatives in L_p-space, with q providing finer control of smoothness. Wavelet bases allow characterization of Besov spaces, with α, p, and q capturing the decay of wavelet coefficients. For further details, see Section F in the Supplementary Material.

We consider the Cohen, Daubechies, and Vial construction of compactly supported, orthonormal, A-regular wavelet basis of $L_2[0,1]$, for $A > \alpha$. The basis functions are given

by $\phi_{l_0+1,m}, \psi_{lk}$ for $m \in \{0, ..., 2^{l_0+1} - 1\}$, $l \geq l_0 + 1$, and $k \in \{0, ..., 2^l - 1\}$, with $\psi_{lk}(x) = 2^{l/2}\psi(2^l x - k)$ and $\phi_{l_0+1,k}(x) = 2^{l_0+1}\phi(2^{l_0+1}x - m)$. For other values of k and m, functions are specially constructed to form a basis with required smoothness, see Section F. Using slight notation abuse, we denote the father wavelet by $\psi_{l_0 k} = \phi_{l_0+1,k}$ and represent any $f \in L_2[0,1]$ as $f = \sum_{l=l_0}^{\infty} \sum_{k=0}^{2^l-1} f_{lk}\psi_{lk}$, where $f_{lk} = \langle f, \psi_{lk} \rangle$ are *wavelet coefficients*. By wavelet orthonormality, $\|f\|_2^2 = \sum_{l=l_0}^{\infty} \sum_{k=0}^{2^l-1} f_{lk}^2$.

We construct our estimators using an A-smooth wavelet basis ($A > \alpha$) with compactly supported ψ such that wavelets $\psi_{lk}(x) = 2^{l/2}\psi(2^l x - k)$ for $l \geq l_0$ and $k = 0, \ldots, 2^l - 1$ form an orthonormal basis for $\mathcal{B}_{p,q}^{\alpha}[0,1]$.

We describe briefly how wavelets are used to construct optimal global and pointwise estimators, based on wavelet approximations up to a limited resolution level. Wavelets' approximation properties in Besov spaces (see e.g. [37]) ensure that changes in data $X_i^{(j)}$ produce limited changes in the wavelet estimator size. The wavelets' limited support shrinks at higher resolution levels, controlling the number of coefficients affected by changes in $X_i^{(j)}$. This ensures individual data changes have limited impact on the shared transcript, which is key to privacy. A detailed description of these properties appears in Sections 2.2 and 2.3.

2.2 Constructing an optimal global estimator

We now construct the estimator using the wavelet transform, which allows representing a function f in L_2 as a linear combination of wavelet basis functions. We first introduce some notation. For $\tau > 0$ and $x \in \mathbb{R}$, let $[x]_\tau$ denote x clipped at the threshold τ: $[x]_\tau := \max(-\tau, \min(\tau, x))$. Given $L \in \mathbb{N}$ and $\tau > 0$, each machine $j = 1, \ldots, m$ computes the real numbers

$$\hat{f}_{lk;\tau}^{(j)} = \frac{1}{n_j} \sum_{i=1}^{n_j} \left[Y_i^{(j)} \right]_\tau \psi_{lk}\left(X_i^{(j)} \right), \tag{6}$$

for $l_0 \leq l \leq L$, $0 \leq k \leq 2^l - 1$. We will specify τ and L later. These numbers form the vector $\hat{\boldsymbol{f}}_{L,\tau}^{(j)} := \left\{ \hat{f}_{lk;\tau}^{(j)} : k = 0, \ldots, 2^l - 1,\ l = l_0, \ldots, L \right\}$, which will underlie our transcript. To ensure privacy, we transmit a noisy version of this vector. Adding noise degrades

the estimator, but sufficient noise ensures the transcript satisfies the privacy guarantee of Definition 1.1. To control the required noise magnitude, it's important that the statistic doesn't change drastically when a single data point changes. We formalize this in terms of the *sensitivity* of the statistic $\hat{\boldsymbol{f}}_{L,\tau}^{(j)}$.

The following lemma bounds the L_1-*sensitivity* of the statistic $\hat{\boldsymbol{f}}_{L,\tau}^{(j)}$, i.e., the difference in L_1 norm when applied to two neighboring datasets.

Lemma 2.1. *For any neighboring datasets $Z^{(j)}$ and $\tilde{Z}^{(j)}$, $\left\|\hat{\boldsymbol{f}}_{L,\tau}^{(j)}(Z^{(j)}) - \hat{\boldsymbol{f}}_{L,\tau}^{(j)}(\tilde{Z}^{(j)})\right\|_1$ is bounded by $c_\psi \frac{\tau\sqrt{2^L}}{n_j}$, where c_ψ depends only on the wavelet basis.*

The proof is provided in Section H.1.1. The limited L_1-sensitivity of $\hat{\boldsymbol{f}}_{L,\tau}^{(j)}$ arises from two factors. The commonly approached strategy of clipping constrains changes in (6) when changing a datum $Y_i^{(j)}$. Secondly, the wavelets have compact support, which shrinks proportionally to when the resolution level l increases. Consequently, eventhough the wavelet basis elements grow exponentially with resolution level l, their support shrinks proportionally. By considering sum of products $\left[Y_i^{(j)}\right]_\tau \psi_{lk}\left(X_i^{(j)}\right)$, this construction greatly limits the number of terms affected in (6) between changes in the i-th datum.

The bounded L_1-sensitivity ensures that $\hat{\boldsymbol{f}}_{L,\tau}^{(j)}(Z^{(j)})$ combined with Laplace noise satisfies $(\varepsilon_j, 0)$-differential privacy. Specifically, the j-th server outputs $T_{lk;\tau}^{(j)} = \hat{f}_{lk;\tau}^{(j)} + W_{lk}^{(j)}$ for $k = 0, \ldots, 2^l - 1$ and $l = l_0, \ldots, L$, where $\boldsymbol{W}^{(j)} := (W_{lk}^{(j)})$ are i.i.d. Laplace with variance $\frac{2\tau^2 2^L c_\psi^2}{n_j^2 \varepsilon_j^2}$. Here, $c_\psi := 4c_A\|\psi\|_\infty$, matching the constant from Lemma 2.1. Thus, the transcript $T_{L,\tau}^{(j)} = \hat{\boldsymbol{f}}_{L,\tau}^{(j)}(Z^{(j)}) + \boldsymbol{W}^{(j)}$ is $(\varepsilon_j, 0)$-differentially private.

The final estimator of f is obtained by carefully reweighting the transcripts, accounting for heterogeneity between servers. The weights depend on the number of local observations n_j and the privacy constraint ε_j. Given the transcripts $T = (T_{L,\tau}^{(1)}, \ldots, T_{L,\tau}^{(m)})$, the final estimator is

$$\hat{f}_{L,\tau}(x) = \sum_{l=l_0}^{L} \sum_{k=0}^{2^l-1} \left(\sum_{j=1}^{m} u_j T_{lk;\tau}^{(j)} \right) \psi_{k,l}(x), \tag{7}$$

with weights

$$u_j = \frac{v_j}{\sum_j v_j} \quad \text{where} \quad v_j = \left(n_j^2 \varepsilon_j^2\right) \wedge \left(n_j 2^L\right). \tag{8}$$

The following theorem captures the global risk attained by the estimator $\hat{f}_{L,\tau}$ resulting from the distributed $(\boldsymbol{\epsilon}, \mathbf{0})$-FDP procedure outlined above, with optimal selection of L and a sufficiently large choice of τ. For the latter, a choice of $C_{\alpha,R} + \sqrt{(2\alpha+1)L}$ is adequate, where $C_{\alpha,R} > 0$ is a constant, as specified by Lemma H.6.

The variance of the Laplace noise vectors $\boldsymbol{W}^{(j)}$, which yield the privacy guarantee, increases with L. Consequently, the optimal choice of L is not just governed by the classical bias variance trade-off, but also by the trade-off in the additional noise required to guarantee privacy. The optimal choice of L is taken as follows. Let $D > 0$ be the number solving the equation

$$D^{2\alpha+2} = \sum_{j=1}^{m} \left(n_j^2 \varepsilon_j^2\right) \wedge \left(n_j D\right). \tag{9}$$

Setting $L = (l_0+1) \vee \lceil \log_2(D) \rceil$ yields the optimal performance as described by the theorem below, in terms of a bias-variance-sensitivity trade-off.

Theorem 2.2. *Set $\tau = C_{\alpha,R} + \sqrt{(2\alpha+1)L}$ and take $L = (l_0+1) \vee \lceil \log_2(D) \rceil$, where $D > 0$ is the solution to (9). Then, the L_2-risk of the distributed $(\boldsymbol{\varepsilon}, \mathbf{0})$-DP protocol $\hat{f}_{L,\tau}$ satisfies*

$$\sup_{f \in \mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f \left\| \hat{f}_{L,\tau} - f \right\|_2^2 \leq C_\psi \log(N) 2^{-2L\alpha},$$

where C_ψ denotes a constant depending on ψ.

We briefly comment on the derived result. It is important to note that a unique positive solution to (9) always exists, as the exponent $2\alpha + 2 > 2$ implies that the left-hand side is smaller than the right-hand side for $D > 0$ small enough, whilst the right-hand side grows linearly for small enough $D > 0$. Furthermore, the right-hand side increases sublinearly in D, whilst the left-hand side increases superlinearly (strictly so).

When the privacy budget is large enough (e.g. $\varepsilon_j = \infty$ for $j = 1, \ldots, m$), D would

be proportional to the number of wavelet coefficients needed to obtain a wavelet estimator that attains the optimal estimation rate, see for example [26]. For $\alpha > 0$ smooth functions in a Besov space, the optimal resolution level of a wavelet estimator would correspond to $(1+2\alpha)^{-1}\lceil \log_2 N \rceil$ for the global risk. However, under privacy constraints, the effective resolution level changes to $(2+2\alpha)^{-1}\lceil \log_2 D \rceil$, which can be substantially different from the case without privacy constraints.

In the next section, it is shown that the performance described above is the theoretically best possible in a minimax sense (up to a logarithmic factor).

2.3 Constructing an optimal estimator of f at a point

We now turn to the task of estimating the unknown function $f \in \mathcal{B}^{\alpha,R}_{p,q}$ at a given point $x_0 \in (0,1)$. That is to say, we will construct an estimator $\hat{f}$ such that $\mathbb{E}_f(\hat{f}(x_0) - f(x_0))^2$ achieves the optimal rates as predicated by Corollaries 2 and 1.

The plug-in estimator from the previous section, $\hat{f}_{L,\tau}(x_0)$, where $\hat{f}_{L,\tau}$ is constructed as described earlier, forms a natural starting point for constructing the pointwise estimator. However, the optimal choice of L for the pointwise risk may differ from the one used to attain the optimal rate for the global risk. As was the case with the estimator of global risk as presented in Section 2.2, there is a trade-off between bias, variance and sensitivity. This trade-off is different in the case of pointwise risk in Besov spaces with $p < \infty$. Here, optimal choice of L is governed by $L = (l_0+1) \vee \lceil \log_2(D) \rceil$, where $D > 0$ be the number solving the equation

$$D^{2\nu+2} = \sum_{j=1}^{m} \left(n_j^2 \varepsilon_j^2\right) \wedge (n_j D)\,. \tag{10}$$

The following theorem describes the performance of the pointwise estimator $\hat{f}_{L,\tau}(x_0)$ on the basis of the $(\boldsymbol{\varepsilon}, \mathbf{0})$-FDP transcript $T = (T^{(1)}_{L,\tau}, \ldots, T^{(m)}_{L,\tau})$ for L chosen as above and $\tau = C_{\nu,R} + \sqrt{2(2\nu+1)L}$, with $C_{\nu,R} > 0$ as given by Lemma H.6.

Theorem 2.3. *Set $\tau = C_{\nu,R} + \sqrt{2(2\nu+1)L}$ and $L = (l_0+1) \vee \lceil \log_2(D) \rceil$, where D is*

governed by (10). *Then, the pointwise risk of the distributed* $(\boldsymbol{\varepsilon}, \mathbf{0})$*-DP protocol* $\hat{f}_{L,\tau}$ *satisfies*

$$\sup_{f \in \mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f(\hat{f}_{L,\tau}(x_0) - f(x_0))^2 \leq C_\psi \log(N) 2^{-2L\nu}.$$

The rate attained by the choice of L as directed by (10) does not just yield the best possible bias-variance-sensitivity trade-off for the estimator class under consideration, but it turns out to be minimax optimal (up to a logarithmic factor) as established in the lower bound of Theorem 3.5 in the following section.

3 Minimax Lower Bounds and Optimality of the Estimators

Theorems 2.2 and 2.3 provide the convergence rates for the proposed estimators of f and $f(x_0)$. In this section, we establish two minimax lower bounds, matching these rates up to logarithmic factors, for both global and pointwise estimation. Together, the upper and lower bounds confirm the minimax rate for FDP estimation, summarized in Theorem 3.1 in Section 3.1. The derivation of each lower bound uses distinct techniques, elaborated in Sections 3.2 and 3.3. For global risk, the technique is reminiscent of the score attack of [21, 22], a generalization of the tracing adversary method [15, 31]. For pointwise risk, we use a coupling argument [6, 43] with Le Cam's two-point method, detailed in Section 3.3. Formal proofs are in Section H.2 of the Supplementary Material.

3.1 Implications of the minimax optimal rates of convergence

In this section, we present our primary findings regarding the minimax rate of convergence under DP constraints. Our results address both the global and pointwise risks.

For the global risk, the minimax rates are encapsulated in the upper bound of Theorem 2.2 and the lower bound of Theorem 3.2, derived in Sections 2.2 and 3.2. Similarly, for

the pointwise risk, our findings are summarized in Theorems 2.3 and 3.5, in the form of an upper bound and lower bound respectively, in Sections 2.3 and 3.3. Together, these theorems are summarized by the following result.

Theorem 3.1. *For $\gamma > 0$, let $D > 0$ be the number solving the equation*

$$D^{2\gamma+2} = \sum_{j=1}^{m} \left(n_j^2 \varepsilon_j^2\right) \wedge (n_j D)\,. \tag{11}$$

Taking $\gamma = \alpha$, the minimax rate for the global risk is given by

$$\inf_{\hat{f} \in \mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})} \sup_{f \in \mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f \|\hat{f} - f\|_2^2 \asymp \left(M_N D^{-2\alpha} \wedge 1\right),$$

whenever for all $j = 1, \ldots, m$ we have $\delta_j \lesssim (n_j^{1/2} \varepsilon_j^2 (D \vee 1)^{-1})^{1+\kappa}$ for some $\kappa > 0$ and where $M_N \geq 1$ is a sequence of the order at most $\log(N) \log(1/\min_{j \in [m]} \delta_j)$.

For $\gamma = \nu$, the minimax rate for the pointwise risk is given by

$$\inf_{\hat{f} \in \mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})} \sup_{f \in \mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f \left|\hat{f}(x_0) - f(x_0)\right|^2 \asymp \left(M_N D^{-2\nu} \wedge 1\right),$$

whenever $\sum_j n_j \delta_j \to 0$, for a sequence $M_N \geq 1$ of the order at most $\log(N)$.

We present several specific cases of Theorem 3.1 through corollaries that encapsulate its various implications. Below, we present the corollaries for the homogeneous setting, where all servers have equal privacy budgets. In Section C in the Supplementary Material, we present the corollaries for various heterogeneous settings, where the servers have different privacy budgets.

Corollary 1. *Suppose that $n_j = n$, $\varepsilon_j = \varepsilon$, $\delta_j = \delta$ for $j = 1, \ldots, m$ and assume that $\delta \lesssim (\varepsilon^2/\sqrt{m})^{1+\kappa}$ for some $\kappa > 0$. Then, the global minimax risk over $\mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$ satisfies (4).*

Whenever $n\varepsilon^2 \ll (mn)^{\frac{1}{2\alpha+1}}$, we have that

$$\inf_{\hat{f}\in\mathcal{M}(\boldsymbol{\varepsilon},\boldsymbol{\delta})} \sup_{f\in\mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f\|\hat{f}-f\|_2^2 \asymp M_{m,n}\,(mn)^{-\frac{2\alpha}{2\alpha+1}} \left(m^{\frac{1}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}} \varepsilon^{-2}\right)^{\frac{2\alpha}{2\alpha+2}},$$

which indicates that the minimax estimation error becomes larger than the unconstrained minimax rate ($(mn)^{-\frac{2\alpha}{2\alpha+1}}$) by a factor of $(m^{\frac{1}{(2\alpha+1)}} n^{-\frac{2\alpha}{2\alpha+1}} \varepsilon^{-2})^{\frac{2\alpha}{2\alpha+2}}$ (ignoring the logarithmic factor). This factor can be seen to capture the cost of privacy in terms of the global risk. A smaller ε results in an increase in minimax estimation error, where larger smoothness exacerbates the increase.

A second observation based on the privacy cost factor is the cost of distributing observations in a privacy setting. Specifically, distributing $N = mn$ observations across m machines becomes more challenging as the N observations are spread over more machines, rather than having more observations on fewer machines. This confirms the common understanding that privacy is easier to retain in larger groups. The relative cost of distributing observations is also related to the smoothness, where greater smoothness further increases the cost. More machines require more noise to compensate for fewer observations, affirming that local differentially private methods perform poorly in multi-observation settings and that applying privacy at the observation level is relatively costly.

Classically, the pointwise risk is known to be subject to different phenomena than the global risk over the Besov spaces [19]. Writing $\nu = \alpha - 1/p$ and assuming $\alpha > 1/p$, it is known that the unconstrained pointwise minimax risk satisfies

$$\inf_{\hat{f}} \sup_{f\in\mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f|\hat{f}(x_0)-f(x_0)|^2 \asymp (mn)^{-\frac{2\nu}{2\nu+1}}. \tag{12}$$

Compared to the unconstrained global risk, this indicates that the estimation error at a point is subject to a fundamentally slower convergence rate than the global estimation minimax rate, where the ℓ_p-norm used to measure the smoothness of the Besov ellipsoid influences the minimax estimation performance. Roughly speaking, the "pointwise" in-

tegrability of the derivatives of the function underlying the data impacts the problem of estimation at a point, whilst the global risk remains unaffected. This effect disappears for Hölder alternatives, where $p = \infty$ and the minimax rate for the global risk and the pointwise risk coincide. The main theorem on the minimax risk for pointwise estimation leads to the following result for the homogeneous setting.

Corollary 2. *Suppose that $n_j = n$, $\varepsilon_j = \varepsilon$, $\delta_j = \delta$ for $j = 1, \dots, m$ and $\delta \ll (mn)^{-1}$. Then, for $x_0 \in [0,1]$, the pointwise minimax risk at x_0 over the class $\mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$ satisfies* (5).

The minimax rate for the pointwise risk seemingly takes on a similar form as that of the global risk and it coincides with the global risk whenever $p = \infty$. However, for finite values of p, the cost of privacy can be seen to differ. In particular, to attain the unconstrained optimal pointwise minimax rate (12), it can be seen that a relatively larger ε is needed, where a smaller value of p in fact exacerbates the demand. More precisely, whenever $(mn)^{\frac{1}{2\alpha+1}} \lesssim n\varepsilon^2 \ll (mn)^{\frac{1}{2\nu+1}}$, the pointwise risk suffers from the DP constraints, whereas the global risk performance is the same as in the problem without the DP constraints.

Whenever $n\varepsilon^2 \ll (mn)^{\frac{1}{2\nu+1}}$, comparing (5) to (12) shows that the minimax rate of the classical (unconstrained) pointwise risk increases by a factor of $(m^{\frac{1}{2\nu+1}} n^{-\frac{2\nu}{2\nu+1}} \varepsilon^{-2})^{\frac{2\nu}{2\nu+2}}$ (ignoring the logarithmic factor). This shows that the pointwise risk is subject to a similar cost-relationship as the global risk. What is similar is that more stringent privacy demands in terms of a smaller ε translate to an increased cost in terms of the pointwise risk. However, the relative increase in privacy cost resulting from a decrease in ε for the case of pointwise risk, is smaller than the relative increase in privacy cost of the global risk, where this discrepancy is further exacarbated for smaller values of p. This shows that stringent privacy demands are comparatively less costly for the pointwise risk.

On the other hand, the cost of distributing observations (i.e. increasing m when distributing $N = nm$ observations) is relatively larger for smaller values of p. That is to say, differentially private estimation in pointwise risk suffers less from stringent per machine privacy demands, while it suffers more from the fact that data is distributed before privacy

preservation is applied. This surprising phenomenon shows that in a distributed setting with privacy constraints, the distribution of the data across servers impacts the rate differently depending on the inferential task at hand. Further results for the heterogeneous setting can be found in Section C in the Supplementary Material.

3.2 Minimax lower bounds for global risk

The following theorem states a lower bound on the minimax risk for global estimation.

Theorem 3.2. *Let $D > 0$ be the solution to (11) and assume that $\delta_j < (n_j^{1/2}\varepsilon_j^2(D \vee 1)^{-1})$ for some $\kappa > 0$ and all $j \in [m]$. Then, we have the following lower bound on the minimax risk:*

$$\inf_{\hat{f} \in \mathcal{M}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})} \sup_{f \in \mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f \|\hat{f} - f\|_2^2 \gtrsim D^{-2\alpha} \wedge 1. \tag{13}$$

Note that this lower bound (up to a log factor) matches the upper bound from Theorem 2.2, for $L = (l_0 + 1) \vee \lceil \log_2(D) \rceil$ in the estimator from Section 2.2, confirming that the proposed estimator attains the best rate among privacy-constrained estimators.

Next, we outline key steps in the proof, with technical details left to the appendix. To lower bound the global risk, we restrict to a finite-dimensional sub-model of $\mathcal{B}_{p,q}^{\alpha,R}$, using the wavelet basis from the previous section. Given $L \in \mathbb{N}$, consider the subspace $\left\{ f \in \mathcal{B}_{p,q}^{\alpha,R} : f = \sum_{k=0}^{2^L-1} f_{Lk}\psi_{Lk},\ f_{Lk} \in [-2^{-L(\alpha+1/2)}R, 2^{-L(\alpha+1/2)}R] \right\}$, denoted by $\mathcal{B}_{p,q}^{\alpha,R,L}$. Let $\boldsymbol{\psi}(X)$ be the 2^L dimensional vector $\{\psi_{Lk}(X)\}_{k=1}^{2^L}$ and define

$$\boldsymbol{S}_f\left(Z_i^{(j)}\right) := \left(Y_i^{(j)} - \sum_{k=0}^{2^L-1} f_{Lk}\psi_{Lk}\left(X_i^{(j)}\right) \right) \boldsymbol{\psi}\left(X_i^{(j)}\right). \tag{14}$$

The random vector $\boldsymbol{S}_f(Z_i^{(j)})$ can be seen as an "score function" of the i-th observation on the $j-$th server, within the finite dimensional sub-model. Similarly, consider the "score function" for local data $Z^{(j)}$ on the jth server; $\bar{\boldsymbol{S}}_f(Z^{(j)}) := \sum_{i=1}^{n_j} \boldsymbol{S}_f(Z_i^{(j)})$. Furthermore, let $\boldsymbol{C}_f^{T^{(j)}}$ denote the 2^L-dimensional random matrix $\mathbb{E}\left[\bar{\boldsymbol{S}}_f\left(Z^{(j)}\right) \mid T^{(j)}\right] \mathbb{E}\left[\bar{\boldsymbol{S}}_f\left(Z^{(j)}\right) \mid T^{(j)}\right]^T$.

We shall write $\boldsymbol{C}_f^{Z^{(j)}}$ for the matrix $\sum_{i=1}^{n_j} \boldsymbol{C}_{f,i}^{(j)}$, where $\boldsymbol{C}_{f,i}^{(j)} = \left[\boldsymbol{S}_f(Z_i^{(j)})\boldsymbol{S}_f(Z_i^{(j)})^T\right]$.

Using the Van-Trees inequality (with a prior as specified later on in the section), we obtain an expression in terms of the sum-of-traces of the matrices $\mathbb{E}\boldsymbol{C}_f^{T^{(j)}}$, i.e. the covariance of the score function $\mathbb{E}\bar{\boldsymbol{S}}_f(Z^{(j)})$, conditionally on the released transcripts.

As the conditional expectation contracts the L_2-norm, we have the "data processing" bound $\mathbb{E}\boldsymbol{C}_f^{T^{(j)}} \leq \mathbb{E}\boldsymbol{C}_f^{Z^{(j)}}$, which in turn implies that $\mathbb{E}\mathrm{Tr}(\boldsymbol{C}_f^{T^{(j)}}) \leq \mathbb{E}\mathrm{Tr}(\boldsymbol{C}_f^{Z^{(j)}})$.

The right-hand side is bounded by $2^L n_j$ by direct calculation, which is detailed in the appendix H.2. These bounds ignore privacy constraints and reflect unconstrained minimax rate. To capture the information loss from the DP constraint in Definition 1.1, a more sophisticated data processing argument is needed. This leads to one of the paper's key innovations: a data-processing inequality (Lemma 3.3) for the conditional covariance given a $(\varepsilon_j, \delta_j)$-differentially private transcript of linear functionals like the score $\boldsymbol{S}_f(Z_i^{(j)})$. The lemma offers a geometric version of the "score attack" lower bound from [22]. Combining this step with trace linearity accommodates heterogeneity between servers.

Lemma 3.3. *Let $\delta_j \log(1/\delta_j) < n_j^{1/2}\varepsilon_j^2(D \vee 1)^{-1}$ for $j = 1, \ldots, m$. There exists a universal constant $C > 0$ such that*

$$\mathbb{E}\mathrm{Tr}(\boldsymbol{C}_f^{T^{(j)}}) \leq C n_j \varepsilon_j \sqrt{\mathbb{E}\mathrm{Tr}(\boldsymbol{C}_f^{T^{(j)}})}\sqrt{\lambda_{\max}(\mathbb{E}\boldsymbol{C}_{f,1}^{(j)})} + C\delta_j\left(2^L n_j^{1/2}\log(1/\delta_j) + n_j\right).$$

In Section H.2 of the appendix, we show that the largest eigenvalue of $\boldsymbol{C}_{f,i}$; $\lambda_{\max}(\mathbb{E}C_{f,i})$, is bounded, from which it follows from the $\mathbb{E}\left[\mathrm{Tr}(\boldsymbol{C}_f^{T^{(j)}})\right] \lesssim n_j^2\varepsilon_j^2$ uniformly for $f \in \mathcal{B}_{p,q}^{\alpha,R,L}$ whenever δ_j is of smaller than $n_j^{1/2}\varepsilon_j^2 D^{-1}$.

With the two bounds on the trace of $\mathbb{E}\boldsymbol{C}_f^{T^{(j)}}$ in hand, we now lower bound global estimation risk using the Van-Trees inequality. The Van-Trees inequality provides an expression in terms of the trace of a certain covariance matrix, which is the conditional covariance of a linear functional of the data. Combined with the data processing inequalities, the linearity of the trace accommodates for the heterogeneity between the servers.

In order to apply the Van-Trees inequality, we first define a prior such that the worst-case global risk is lower bounded by the corresponding Bayes risk. To that extent, we define a prior Π that is supported on $\mathcal{B}_{p,q}^{\alpha,R,L}$. Given the resolution level $L \in \mathbb{N}$, we draw f_{Lk} independently from the probability distribution Π_{Lk}, defined through an appropriately rescaled version of the density $t \mapsto \cos^2(\pi t/2)\mathbb{1}_{|t|\leq 1)}$ such that has its support equal to $[-2^{-L(\alpha+1/2)}R, 2^{-L(\alpha+1/2)}R]$ for $k = 0\ldots, 2^L - 1$ and set $f_{lk} = 0$ otherwise. For this choice of prior, the Van-Trees inequality of [36] yields the following lemma, for which we defer the details of the proof to Section H.2 in the appendix.

Lemma 3.4. *It holds that* $\sup_{f\in\mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f\|\hat{f}-f\|_2^2$ *is lower bounded by the Bayes risk* $\int \mathbb{E}_f\|\hat{f}-f\|_2^2 d\Pi(f)$*, which is further lower bounded as follows*

$$\int \mathbb{E}_f\|\hat{f} - f\|_2^2 d\Pi(f) \geq \frac{2^{2L}}{\sup_{f\in\mathcal{B}_{p,q}^{\alpha,R,L}} \sum_{j=1}^m \mathbb{E}\mathrm{Tr}(\boldsymbol{C}_f^{T^{(j)}}) + \pi^2 2^{L(2\alpha+2)}}.$$

Combining the data processing upper bound on the trace of $\mathbb{E}\boldsymbol{C}_f^{T^{(j)}}$ and Lemma 3.3, we have, by Lemma 3.4, that

$$\sup_{f\in\mathcal{B}_{p,q}^{\alpha,R}} \mathbb{E}_f\|\hat{f} - f\|_2^2 \gtrsim \frac{2^{2L}}{\sum_{j=1}^m n_j^2\varepsilon_j^2 \wedge n_j 2^L + \pi^2 2^{L(2\alpha+2)}}.$$

For obtaining the desired lower bound we choose L that maximizes the lower bound. Setting $L = (l_0 + 1) \vee \lceil \log_2(D) \rceil$ does so by the relationship (11), which proves Theorem 3.2.

3.3 Lower bound for the pointwise risk

In this section, we derive the minimax lower bound for the pointwise risk. We first present the lower bound as the main result of the section in the form of Theorem 3.5, after which we discuss its proof. The theorem tells us that the pointwise risk estimator proposed in Section 2.3 performs optimally in terms of achieving the minimax privacy constrained rate up to a logarithmic factor.

Theorem 3.5. *Assume furthermore that* $\sum_j n_j\delta_j \to 0$ *and let* $D > 0$ *be the number solving the equation*

$$D^{2\nu+2} = \sum_{j=1}^{m} \left(n_j^2\varepsilon_j^2\right) \wedge (n_j D)\,. \tag{15}$$

Then, for any $x_0 \in (0,1)$*, the minimax pointwise risk is lower bounded as follows:*

$$\inf_{\hat{f}\in\mathcal{M}(\boldsymbol{\varepsilon},\boldsymbol{\delta})} \sup_{f\in\mathcal{B}^{\alpha,R}_{p,q}} \mathbb{E}_f|\hat{f}(x_0) - f(x_0)|^2 \gtrsim D^{-2\nu} \wedge 1.$$

Whenever $\left(\sum_{j=1}^m n_j^2\varepsilon_j^2\right)^{\frac{1}{2\nu+2}} \geq \max_j n_j\epsilon_j^2$*, the right hand side is further bounded from below by* $\left(\sum_{j=1}^m n_j^2\varepsilon_j^2\right)^{-\frac{2\nu}{2\nu+2}} \wedge 1$.

The proof of the theorem is based around the Le Cam two point method, which is a common approach to lower bounding the pointwise risk. However, to capture the effect of the transcripts satisfying the DP constraint of Definition 1.1, we introduce a coupling argument in conjunction.

We briefly sketch the two point method and coupling argument here, leaving the technical details to the appendix. Take any function $f \in \mathcal{B}^{\alpha}_{p,q}$ such that $\|f\|_{\mathcal{B}^{\alpha}_{p,q}} = R' < R$ and a compactly supported function $g \in \mathcal{B}^{\alpha}_{p,q}$ such that $\|g\|_{\mathcal{B}^{\alpha}_{p,q}} \leq R - R'$ and $g(0) > 0$. Define a third function $\tilde{f}(t) := \gamma_D^{-1} g(\beta_D(t - x_0)) + f(t)$, where $\gamma_D := c_0^{-1} D^{\nu}$ and $\beta_D = \gamma_D^{1/\nu}$, where we recall that $\nu = \alpha - \frac{1}{p}$. By e.g. Lemma 1 from [19], $\|f\|_{\mathcal{B}^{\alpha}_{p,q}} \leq R$.

Let $(Y_i^{(j)}, X_i^{(j)}) \sim P_f$ and $(\tilde{Y}_i^{(j)}, \tilde{X}_i^{(j)}) \sim P_{\tilde{f}}$ for individual observations generated according to (1) with either f or $\tilde{f}$ the true underlying regression function respectively. We construct a coupling between P_f and $P_{\tilde{f}}$ such that $(Y_i^{(j)}, X_i^{(j)})$ and $(\tilde{Y}_i^{(j)}, \tilde{X}_i^{(j)})$ are equal with probability proportional to $\sigma^{-1}\|\tilde{f} - f\|_1$, which forms the content of the following lemma.

Lemma 3.6. *There exists a joint distribution* $P_{f,\tilde{f}}$ *of* $\left((Y_i^{(j)}, X_i^{(j)}), (\tilde{Y}_i^{(j)}, \tilde{X}_i^{(j)})\right)$ *such that*

$$\rho := P_{f,\tilde{f}}\left(\left(Y_i^{(j)}, X_i^{(j)}\right) \neq \left(\tilde{Y}_i^{(j)}, \tilde{X}_i^{(j)}\right)\right) \leq \frac{c}{\sigma}\|\tilde{f} - f\|_1, \tag{16}$$

for a universal constant $c > 0$.

We prove the above lemma in Section I.2. Loosely speaking, the quantity ρ captures the difficulty of distinguishing individual observations from P_f of those generated from $P_{\tilde{f}}$.

Consider now transcripts $T = (T^{(1)}, \ldots, T^{(m)})$ each satisfying the DP constraint of Definition 1.1 with a privacy budget $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$, and let $\mathbb{P}_f$ denote the joint law of transcripts and the $N = \sum_{j=1}^{m} n_j$ observations generated from P_f. Let $\mathbb{P}_f^T$ denote the push-forward measure of the transcript, i.e. its marginal distribution given that the data is generated by P_f. Similarly, let $\mathbb{P}_{\tilde{f}}$ denote the joint law of T with the data generated from $P_{\tilde{f}}$ and let $\mathbb{P}_{\tilde{f}}^T$ denote the corresponding marginal distribution of T. With the coupling of Lemma 3.6 in hand, we derive the following lemma.

Lemma 3.7. *For any subset $S \subseteq [m]$, with $\bar{\varepsilon}_j = 6 n_j \varepsilon_j \rho$, ρ as defined in* (16)*, it holds that*

$$\left\| \mathbb{P}_f^T - \mathbb{P}_{\tilde{f}}^T \right\|_{\mathrm{TV}} \leq \sqrt{2} \sqrt{\sum_{j \in S} \bar{\varepsilon}_j \left(e^{\bar{\varepsilon}_j} - 1\right) + \sum_{j \in S^c} n_j D_{\mathrm{KL}}(P_f; P_{\tilde{f}})} + 4 \sum_{j \in S} e^{\bar{\varepsilon}_j} n_j \delta_j \rho, \tag{17}$$

We defer a proof of the lemma to Section I.2 of the appendix. The lemma allows analysis of the contributions of the separate the servers, accounting for the heterogeneity in the privacy budgets (ϵ_j, δ_j) and the differing number of observations. Roughly speaking, for servers with relatively large privacy budgets, their contribution to the estimator is to be captured by $n_j D_{\mathrm{KL}}(P_f; P_{\tilde{f}})$, which does not involve the privacy budget all together. Servers for which the privacy budget is more stringent, contribute with the (potentially) smaller quantity $\bar{\varepsilon}_j$, where ρ corresponds to the probability in (16), established in the coupling relationship of Lemma 3.6. The proof then follows by Le Cam's two point lemma after combining the divergence bound of Lemma 3.6 and a standard KL-divergence bound and minimizing the right-hand side through the choice of S, i.e. the optimal division into the stringent and non-stringent privacy budgets, we defer the details to Section I.2 of the Supplementary Material.

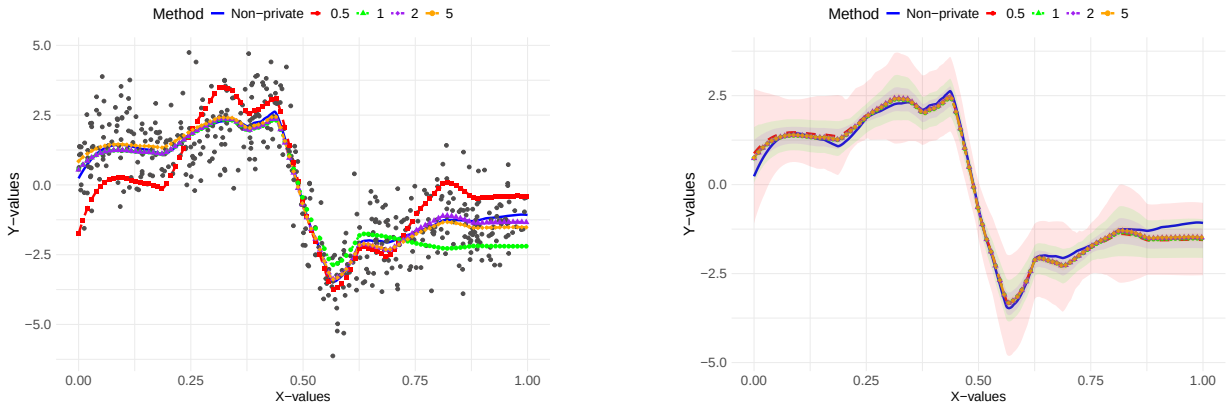
4 Simulation Results and Real Data Applications

4.1 Simulation Studies

In this section, we present numerical results related to the estimator constructed in Section 2. The purpose of our simulations is two-fold. First of all, we investigate how much the noise added by the differential privacy mechanism affects the the estimator, by comparing various privacy budgets. Secondly, we illustrate numerically how the estimation risk is affected by the number of servers, the total number of observations, and the privacy budget.

As an initial investigation into the effect of the privacy mechanism on the estimator, we consider a simple example with a single server and a total of $N = 500$ observations in Figure 2. Here, we generate data from the model (1), where f is randomly drawn from a Besov ball; we defer the exact details of the simulation setup to Section E.1 in the Supplementary Material. The plot in Figure 2a shows a single draw the private estimator of Section 2 is compared for different privacy budgets ($\epsilon = 0.5, 1, 2, 5$), the non-private wavelet estimator and the true underlying signal. The plot in Figure 2b shows the mean estimated signals over 1000 simulation runs, with bands capturing 95% of the draws over the randomness in the privacy mechanism. We observe that (as expected) greater deviances from the true signal / non-private estimator occuring for smaller privacy budgets.

Figure 3 contains plots studying how the the MSE changes for various characteristics of the problem. In the plot Figure 3a, we show the effect of increasing the privacy budget on the IMSE for different server setups, for a fixed number of observations ($N = 2000$). We see that for small values of ϵ the performance is better when the data is distributed over fewer machines. As ϵ increases, the performance becomes independent of the number of machines. This captures the phase transition also observed in our theoretical results. The two slight increases in IMSE observed in the plot are a consequence of the changing resolution levels of the wavelet expansion that occur as ε increases, where we have not optimized the constants when selecting the change points. The second plot, Figure 3b,



(a) Scatter points and estimated curves for different privacy budgets.

(b) Mean and 95% bands over 1000 runs over the privacy mechanism.

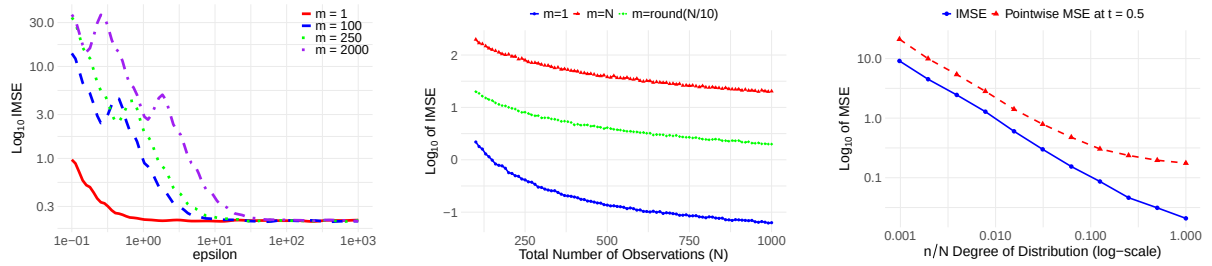
Figure 2: Wavelet-based differentially private estimators for varying privacy budgets ($\epsilon = 0.5, 1, 2, 5$) with a total of $N = 500$ observations on one single server. The left plot shows the scatter points and a single estimated curve for each value of ϵ, while the right plot compares the mean estimated signals over 1000 simulation runs over the randomness in privacy mechanism with bands capturing 95% of the draws over the privacy mechanism.

shows the decrease in the IMSE as N increases, for a fixed privacy budget ($\epsilon = 1$) on a log-log scale. We see that the performance improves as the number of observations increases, as expected. However, we see that the slope of the curve is steeper for the case of a single machine compared to the case of multiple machines, which is consistent with the theoretical results. The third plot, Figure 3c, further corroborates the findings concerning the cost of distributing data over multiple servers by showing the effect of decreasing the number of servers ($n/N \to 1$) for a fixed total number of observations ($N = 2000$), for both IMSE and MSE at a point over.

In Section E.1 of the Supplementary Material, we provide additional simulation results, including tables and figures, that further illustrate the performance of the proposed estimator under various settings.

4.2 Real Data Applications

This section demonstrates the private wavelet based estimator for preserving differential privacy while analyzing relationships between two variables on real data. We consider two



(a) Phase transition ε increases for IMSE.

(b) Decay in IMSE as N increases.

(c) Decreasing the number of servers ($n/N \to 1$).

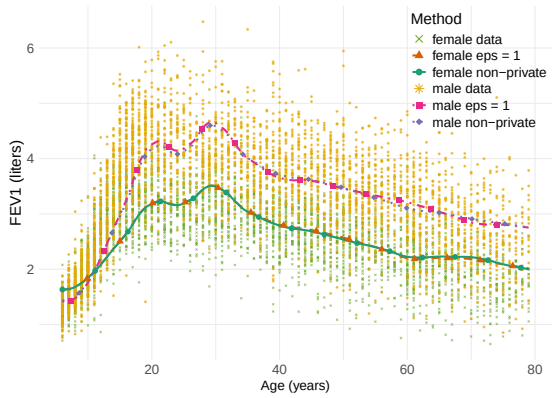
Figure 3: The log-IMSE as a function of the privacy budgets ϵ and total number of observations N, for various machine setups and $\alpha = 2$. The lines represent averages over 1000 simulation runs.

applications: the first example involves the bivariate association between Forced Expiratory Volume in 1 second (FEV1), a measure assessing lung function and respiratory health and age, which is known to follow a non-linear relationship [39]. The second example examines the association between serum folate and serum vitamin B12 levels with plasma total homocysteine (tHcy) levels. Elevated tHcy is a well-established risk factor for vascular diseases, including stroke and myocardial infarction [56] and its relationship with folate and vitamin B12 is characterized by non-linearity [52].

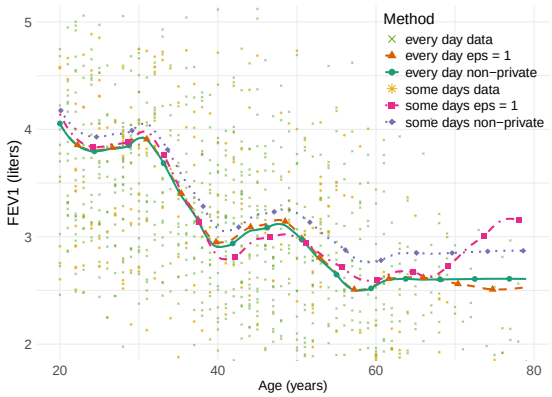
Both applications are based on data of the National Health and Nutrition Examination Survey (NHANES), is available at CDC's NHANES website[2]. We explore the private wavelet-based estimator on features of this otherwise publicly available data for the purpose of illustration. The data consists of a random sample of the U.S. population living in households, selected from 81 counties across the United States.

Returning to the first application, we look at NHANES 2009-2010 data, investigating the relationship between FEV1 and age between ages 6 and 80 years for men and women. In addition, we consider the sub-samples of the survey participants who indicated smoking status. Both gender and smoking status are known to affect FEV1 levels (see e.g. [13] and [49]). Figure 4a displays the estimated curves for smoothness level $\alpha = 2$ – both the private wavelet-based estimator for $\epsilon = 1$ and its non-private counterpart – for women and

[2]https://wwwn.cdc.gov/Nchs/Nhanes



(a) FEV1 vs Age for male and female cohorts.



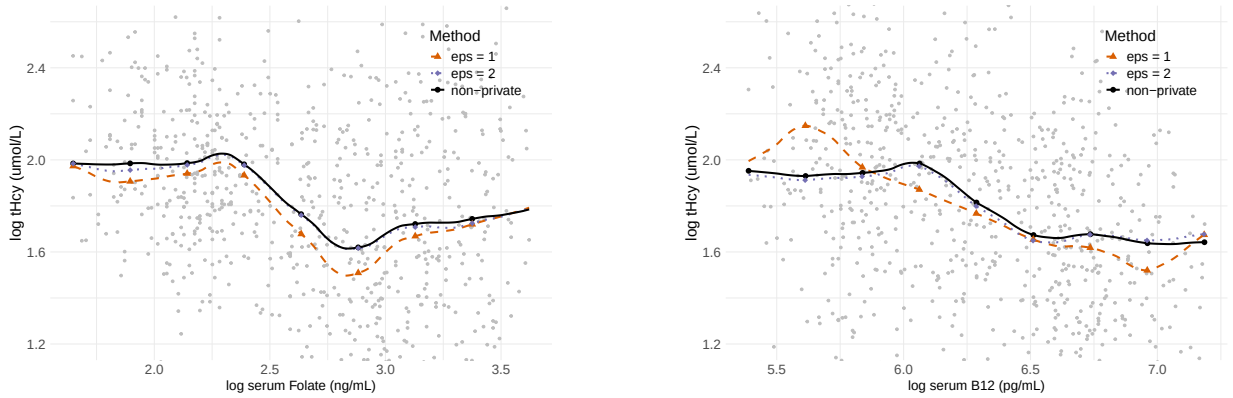
(b) FEV1 vs Age per smoking status.

Figure 4: The relationship between FEV1 and age for female and male cohorts ($n = 3531$ and $n = 3548$, respectively) and participants indicating to smoke every day ($n = 899$) versus those who smoke sometimes ($n = 205$).

men; $n = 3531$ and $n = 3548$, respectively. The curves corroborate the general finding that FEV1 increases with age up until around the 20th year, and declines subsequently, with a noticable difference between the male and female cohorts past the teenage years. In these relatively large samples, the private estimator almost matches the non-private estimator. In Figure 4b, we look at a smaller cohort: individuals ages between 20 and 80 who have indicated to either smoke every day ($n = 899$) or sometimes ($n = 205$). The non-private curves show that the FEV1 levels are lower for individuals who smoke every day compared to those who smoke sometimes. However, the private wavelet-based estimator is less precise in capturing this difference, and could in this particular case lead to an erronous conclusion. This is a consequence of the small sample size of the second cohort, requiring more noise to ensure privacy.

In our second example, we utilize data from the NHANES during 1999-2000, previously studied by Bang et al. [9], who employed quartile plots, piecewise constant models, and splines to illustrate these dose-response curves and reported non-linear inverse relationships between plasma total homocysteine and serum folate and vitamin B12 levels, corroborating general findings on the relationship.

Figure 5 compares the private and non-private dose-response curves between plasma



(a) Folate vs Homocysteine

(b) B12 vs Homocysteine

Figure 5: Private and nonprivate estimation of Homocysteine vs. Folate and B12 for $N = 750$ and $\epsilon \in \{1, 2\}$.

total homocysteine and serum folate and vitamin B12 levels for $N = 750$ and privacy budgets $\epsilon \in \{1, 2\}$. For $\epsilon = 1$, the privacy mechanism introduces a noticeable deviation from the non-private estimator. The plots demonstrate an inverse non-linear relationship between homocysteine, folate, and vitamin B12 levels, consistent with [9]. Figures S.3 and S.4 in Section E provide comparisons for the same privacy budgets with sub-samples of sizes $N = 1500$ and $N = 300$. As sample sizes decrease, the deviation from the non-private estimator increases, in line with our theoretical findings. Full implementation details and further results are provided in the Supplementary Material.

References

[1] Supplement to "Optimal federated learning for nonparametric regression with heterogenous distributed differential privacy constraints".

[2] J. M. Abowd, I. M. Rodriguez, W. N. Sexton, P. E. Singer, and L. Vilhuber. The modernization of statistical disclosure limitation at the us census bureau. *US Census Bureau*, 2020.

[3] J. Acharya, K. Bonawitz, P. Kairouz, D. Ramage, and Z. Sun. Context aware local differential privacy. In H. D. III and A. Singh, editors, *Proceedings of the 37th Inter-*

national Conference on Machine Learning, volume 119, pages 52–62. PMLR, 2020.

[4] J. Acharya, C. L. Canonne, A. V. Singh, and H. Tyagi. Optimal rates for nonparametric density estimation under communication constraints. *IEEE Transactions on Information Theory*, 2023.

[5] J. Acharya, C. L. Canonne, Z. Sun, and H. Tyagi. Unified lower bounds for interactive high-dimensional estimation under information constraints. *Advances in Neural Information Processing Systems*, 36, 2024.

[6] J. Acharya, Y. Liu, and Z. 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.

[7] J. Acharya, Z. Sun, and H. Zhang. Differentially private testing of identity and closeness of discrete distributions. In *Proceedings of the 31st Advances in Neural Information Processing Systems*, volume 31, 2018.

[8] J. Acharya, Z. Sun, and H. Zhang. Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory*, pages 48–78. PMLR, 2021.

[9] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman. Local differential privacy for deep learning. *IEEE Internet of Things Journal*, 7(7):5827–5842, 2019.

[10] H. Bang, M. Mazumdar, and D. Spence. Tutorial in biostatistics: Analyzing associations between total plasma homocysteine and b vitamins using optimal categorization and segmented regression. *Neuroepidemiology*, 27(4):188–200, 2006.

[11] L. P. Barnes, W.-N. Chen, and A. Özgür. Fisher information under local differential privacy. *IEEE Journal on Selected Areas in Information Theory*, 1(3):645–659, 2020.

[12] R. Bassily, A. Smith, and A. Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th annual symposium on foundations of computer science*, pages 464–473. IEEE, 2014.

[13] F. Beaufays, K. Rao, R. Mathews, and S. Ramaswamy. Federated learning for emoji

prediction in a mobile keyboard. *arXiv:1906.04329*, 2019.

[14] M. R. Becklake and F. Kauffmann. Gender differences in airway behaviour over the human life span. *Thorax*, 54(12):1119–1138, 1999.

[15] T. B. Berrett, L. Györfi, and H. Walk. Strongly universally consistent nonparametric regression and classification with privatised data. *Electronic Journal of Statistics*, 2021.

[16] M. Bun, J. Ullman, and S. Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 1–10, 2014.

[17] C. Butucea, A. Dubois, M. Kroll, and A. Saumard. Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli*, 26(3):1727 – 1764, 2020.

[18] T. T. Cai. Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *The Annals of Statistics*, 27(3):898–924, 1999.

[19] T. T. Cai. On block thresholding in wavelet regression: Adaptivity, block size, and threshold level. *Statistica Sinica*, 12:1241–1274, 2002.

[20] T. T. Cai. Rates of convergence and adaptation over Besov spaces under pointwise risk. *Statistica Sinica*, pages 881–902, 2003.

[21] T. T. Cai, Y. Wang, and L. Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics*, 49(5):2825–2850, 2021.

[22] T. T. Cai, Y. Wang, and L. Zhang. Score attack: A lower bound technique for optimal differentially private learning. *arXiv preprint arXiv:2303.07152*, 2023.

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

[24] S. Chaudhuri and T. A. Courtade. Mean estimation under heterogeneous privacy:

Some privacy can be free. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 1639–1644. IEEE, 2023.

[25] B. Ding, J. Kulkarni, and S. Yekhanin. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30, 2017.

[26] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *The annals of Statistics*, 26(3):879–921, 1998.

[27] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.

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

[29] C. Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.

[30] C. Dwork, A. Smith, T. Steinke, and J. Ullman. Exposed! a survey of attacks on private data. *Annual Review of Statistics and Its Application*, 4:61–84, 2017.

[31] C. Dwork, A. Smith, T. Steinke, J. Ullman, and S. Vadhan. Robust traceability from trace amounts. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 650–669. IEEE, 2015.

[32] C. Dwork, K. Talwar, A. Thakurta, and L. Zhang. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 11–20, 2014.

[33] U. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, page 1054–1067, 2014.

[34] A. Fallah, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Optimal and differentially private data acquisition: Central and local mechanisms. *Operations Research*, 2023.

[35] F. Farokhi. Deconvoluting kernel density estimation and regression for locally differentially private data. *Scientific Reports*, 10(1):21361, 2020.

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

[37] E. Gine and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, Cambridge, 2016.

[38] P. Hall, G. Kerkyacharian, and D. Picard. On the minimax optimality of block thresholded wavelet estimators. *Statistica Sinica*, pages 33–49, 1999.

[39] J. L. Hankinson, J. R. Odencrantz, and K. B. Fedan. Spirometric reference values from a sample of the general us population. *American journal of respiratory and critical care medicine*, 159(1):179–187, 1999.

[40] I. Ibragimov and R. Khasminskii. *Some Estimation Problems in Infinite Dimensional Gaussian White Noise*, pages 259–274. Springer New York, 1997.

[41] G. Kamath, J. Li, V. Singhal, and J. Ullman. Privately learning high-dimensional distributions. In *Conference on Learning Theory*, pages 1853–1902. PMLR, 2019.

[42] G. Kamath, V. Singhal, and J. Ullman. Private mean estimation of heavy-tailed distributions. In *Conference on Learning Theory*, pages 2204–2235. PMLR, 2020.

[43] V. Karwa and S. Vadhan. Finite sample differentially private confidence intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, 2018.

[44] M. Kroll. On density estimation at a fixed point under local differential privacy. *Electronic Journal of Statistics*, 15(1):1783 – 1813, 2021.

[45] C. Lalanne, A. Garivier, and R. Gribonval. About the cost of central privacy in density estimation. *Transactions on Machine Learning Research Journal*, 2023.

[46] C. Lalanne, A. Garivier, and R. Gribonval. On the statistical complexity of estimation and testing under privacy constraints. *Transactions on Machine Learning Research Journal*, 2023.

[47] D. Levy, Z. Sun, K. Amin, S. Kale, A. Kulesza, M. Mohri, and A. T. Suresh. Learning

with user-level privacy. *Advances in Neural Information Processing Systems*, 34:12466–12479, 2021.

[48] Y. Liu, A. T. Suresh, F. X. X. Yu, S. Kumar, and M. Riley. Learning discrete distributions: user vs item-level privacy. *Advances in Neural Information Processing Systems*, 33:20965–20976, 2020.

[49] D. M. Mannino and K. J. Davis. Lung function decline and outcomes in an elderly population. *Thorax*, 61(6):472–477, 2006.

[50] S. Narayanan. Private high-dimensional hypothesis testing. In *Conference on Learning Theory*, pages 3979–4027. PMLR, 2022.

[51] S. Narayanan, V. Mirrokni, and H. Esfandiari. Tight and robust private mean estimation with few users. In *International Conference on Machine Learning*, pages 16383–16412. PMLR, 2022.

[52] J. Robertson, F. Iemolo, S. P. Stabler, R. H. Allen, and J. D. Spence. Vitamin b12, homocysteine and carotid plaque in the era of folic acid fortification of enriched cereal grain products. *Cmaj*, 172(12):1569–1573, 2005.

[53] M. Sart. Density estimation under local differential privacy and Hellinger loss. *Bernoulli*, 29(3):2318 – 2341, 2023.

[54] A. Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 813–822, 2011.

[55] A. Team et al. Learning with privacy at scale. *Apple Mach. Learn. J*, 1(8):1–25, 2017.

[56] O. To, D. Case, and S. Nineteen. Plasma homocysteine as a risk factor for vascular disease. *Jama*, 277:1775–1781, 1997.

[57] M. Ye and A. Barg. Optimal schemes for discrete distribution estimation under locally differential privacy. *IEEE Transactions on Information Theory*, 64(8):5662–5676, 2018.