axioms

MDPI

Article

Privacy-Preserving Statistical Inference for Stochastic Frontier Analysis

Mengxiang Quan [1], Yunquan Song [1] and Xinmin Wang [2,*]

[1] College of Science, China University of Petroleum, Qingdao 266580, China; 2202010825@s.upc.edu.cn (M.Q.); syqfly1980@upc.edu.cn (Y.S.)
[2] School of Economics and Management, China University of Petroleum, Qingdao 266580, China
* Correspondence: wangximin@upc.edu.cn

Abstract

We present the first differentially private framework for stochastic frontier analysis (SFA), addressing the challenge of non-convex objectives in privacy-preserving efficiency estimation. We construct a bounded parameter space to control gradient sensitivity and adapt the Frank–Wolfe algorithm with calibrated linear oracle noise to mitigate cumulative perturbation. Incorporating ℓ_1-regularization facilitates sparse and interpretable variable selection under strict (ϵ, δ)-differential privacy. Experiments demonstrate 15–35% MAE reduction under $\epsilon = 0.1$, along with strong scalability and estimation accuracy compared to prior DP methods for non-convex models.

Keywords: stochastic frontier analysis; differential privacy; variable selection; high dimension

MSC: 62J07; 62P20

1. Introduction

Stochastic frontier analysis (SFA) [1] is a widely used method for efficiency measurement that accounts for the difference between actual output and an efficient "frontier" by modeling technical inefficiency and random noise [2,3]. Traditionally, SFA employs parametric models [4,5] estimated via maximum likelihood or Bayesian methods [6]. Recent advances leverage machine learning to capture non-linear relationships, e.g., Kumbhakar [7] uses neural networks for flexible frontier estimation, Fan et al. [8] applies support vector regression to model production functions, and Tsionas [9] combines Bayesian deep learning with SFA for dynamic efficiency analysis.

However, modern SFA applications increasingly involve highly sensitive datasets—patient medical records, proprietary financial data, and confidential agricultural information—that pose severe privacy risks when conducting statistical inference or variable selection in high-dimensional settings [10]. This creates a fundamental challenge, i.e., how to perform accurate efficiency analysis while protecting individual privacy through rigorous mathematical processes.

Differential privacy (DP) [11] is the standard framework for rigorous privacy protection. While DP has been successfully applied to linear and logistic regression, as well as deep learning [12,13], its application to SFA is fundamentally challenging. The main difficulty lies in the non-convexity of the SFA log-likelihood and the complex sensitivity of its efficiency parameters, which violate the convexity assumptions required by most existing DP optimization methods.

check for updates

Academic Editors: Tomas Ruzgas and Angel Ricardo Plastino

Received: 5 June 2025
Revised: 20 August 2025
Accepted: 27 August 2025
Published: 29 August 2025

Citation: Quan, M.; Song, Y.; Wang, X. Privacy-Preserving Statistical Inference for Stochastic Frontier Analysis. *Axioms* **2025**, *14*, 667. https://doi.org/10.3390/axioms14090667

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

For example, DP-stochastic gradient descent (DP-SGD) [13]—a widely adopted approach—adds Gaussian noise to clipped gradients at each iteration. While effective in convex or moderately non-convex problems, DP-SGD becomes unstable for SFA due to cumulative noise and gradient clipping interactions that undermine convergence. Other methods, such as output perturbation [14] or Frank–Wolfe-based mechanisms [15], introduce less noise but require substantial adaptation for non-convex objectives, which have not been developed for SFA.

Recent studies exploring DP alternatives, including one-step mechanisms [16,17] for convex empirical risk minimization and differentially private variance reduction methods [18], remain fundamentally limited by convexity assumptions that make them inapplicable to SFA. Non-gradient DP methods such as Bayesian DP [19] can theoretically handle non-linear efficiency parameters but are computationally prohibitive for large-scale applications. DP-LASSO methods for convex problems [20–22] introduce excessive noise when adapted to non-convex objectives, completely compromising accuracy and variable selection performance.

The Frank–Wolfe (FW) algorithm [23], originally developed for constrained convex optimization, offers theoretical promise due to its projection-free nature and ability to provide sparse solutions [24]. DP-FW methods enhance this framework by selectively adding calibrated noise to linear objectives [25–27]. Acharya et al. [28] introduce personalized DP for ridge regression, but their focus on ℓ_2-regularization does not address sparsity or non-convexity. However, these methods cannot be directly applied to SFA due to the non-convexity of its objective function; no existing work has analyzed the sensitivity properties of SFA's efficiency parameters or developed appropriate noise calibration for the complex parameter interactions inherent in frontier models. Non-convex DP optimization methods, such as DP-projected gradient descent and DP-FW [29,30], provide frameworks for complex losses, but their application to SFA's unique efficiency decomposition and sparse variable selection remains completely unaddressed.

The fundamental gap is that no existing method adequately addresses the intersection of non-convexity, efficiency parameter estimation, sparsity, and privacy, which defines the SFA privacy problem. Existing DP-SGD-based approaches fail due to instability and noise accumulation, while DP-Bayesian and convex DP-LASSO methods are either computationally intensive or fundamentally ill-suited for sparse variable selection under privacy constraints in non-convex problems. To address these critical limitations, we develop the first differentially private optimization framework specifically designed for stochastic frontier models. Our approach is a novel paradigm that combines a constrained Frank–Wolfe optimization framework with LASSO regularization to induce sparsity. This allows us to operate within a carefully constructed bounded convex parameter space that, for the first time, enables us to bound gradient sensitivity for SFA's non-convex parameters. By adding calibrated Laplace noise at the linear oracle level, we fundamentally mitigate the noise accumulation that plagues gradient-based methods like DP-SGD. This provides rigorous (ϵ, δ)-differential privacy guarantees while preserving accuracy and interpretability in high dimensions—capabilities that existing methods cannot achieve.

The key innovations of this paper are as follows:

(1) For non-convex stochastic frontier models, we establish a differentially private estimation framework by deriving bounded gradient sensitivity.
(2) Based on this foundation, we design a constrained Frank–Wolfe optimization algorithm that enables efficient privacy-preserving estimation with calibrated noise.
(3) For high-dimensional settings, we integrate ℓ_1-regularization into the framework to achieve sparse and interpretable variable selection under differential privacy.

The rest of the paper is organized as follows. Section 2 reviews the stochastic frontier model and ℓ_1-regularized variable selection, as well as introducing our bounded parameter space, gradient sensitivity bounds, noise calibration strategy, and a Frank–Wolfe algorithm with an (ϵ, δ)-privacy guarantee. Section 3 presents numerical results, including simulations and real-world validation on the California Housing and FADN datasets. Section 4 concludes the paper and outlines future research directions.

2. Methodology

This section develops our differentially private framework for SFA. It begins with a concise overview of the necessary background concepts before presenting our core methodological contributions—the design of a novel constrained parameter space, a formal sensitivity analysis, and a custom Frank–Wolfe algorithm with rigorous privacy guarantees.

2.1. Stochastic Frontier Model and Maximum Likelihood Estimation

The stochastic frontier model (SFM) assumes that the production process is influenced by both random noise and technical inefficiency. For N production units, the model can be expressed as follows:

$$Y = X\beta + \varepsilon, \quad \varepsilon = \nu - u, \tag{1}$$

where $Y \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times p}$, and $\beta \in \mathbb{R}^p$ are the production frontier parameters.

$\nu \sim \mathcal{N}(0, \sigma_\nu^2 I_N)$ is the symmetric random error, capturing measurement errors and other random factors, where I_N is the N-dimensional identity matrix. $u \sim \mathcal{N}^+(0, \sigma_u^2 I_N)$ is the one-sided error term, representing technical inefficiency, following a zero-truncated half-normal distribution.

We assume ν and u to be uncorrelated and independent of covariates X. The model's parameter set is $\theta = \{\beta, \sigma_\nu^2, \sigma_u^2\}$. For estimation convenience, the total variance is defined as $\sigma^2 = \sigma_\nu^2 + \sigma_u^2$, while the proportion of inefficiency's relative contribution is defined as $\lambda_0 = \sigma_u/\sigma_\nu$.

Common approaches to estimating the stochastic frontier model include corrected least squares, the generalized method of moments, maximum likelihood estimation, and Bayesian methods. Among these, MLE stands out for its simplicity and efficiency. Below, we describe the MLE for the SFM, which forms the foundation for the variable selection techniques that are subsequently addressed.

The probability density function of the composite error term $\epsilon_i = v_i - u_i$ is obtained through convolution techniques. Given the distributional properties of v_i and u_i, the density of ϵ_i incorporates the standard normal cumulative distribution function $\Psi(\cdot)$. The log-likelihood function for the model is expressed as follows:

$$\ln \mathcal{L}(\lambda, \beta, \sigma^2 | y, X) = -\frac{N}{2}\ln\sigma^2 + \sum_{i=1}^{N} \ln \Psi\left(-\frac{\epsilon_i \lambda}{\sigma}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\epsilon_i^2, \tag{2}$$

where $\epsilon_i = y_i - x_i^\top \beta$ represents the residual for the i-th observation, and $\Psi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. MLE seeks to estimate the θ parameters by maximizing the log-likelihood function $\ln \mathcal{L}(\lambda, \beta, \sigma^2 | y, X)$.

2.2. Variable Selection via LASSO and Adaptive LASSO

In practical applications, when the number of covariates p is substantial compared to the sample size N (i.e., high-dimensional data scenarios), traditional MLE tends to cause overfitting and reduce model interpretability. Therefore, we introduce L1 regularization to identify the most explanatory features for output through variable selection, enhancing model robustness and practicality.

The penalized log-likelihood is as follows:

$$L_1(\theta; r) = \ln L(\lambda_0, \beta, \sigma^2) + r \sum_{j=1}^{p} |\beta_j|, \tag{3}$$

where $\theta = \{\lambda_0, \sigma^2, \beta\}$, and r controls the sparsity level. The estimator $\tilde{\theta} = \arg\min_\theta L_1(\theta; r)$ is obtained via gradient-based optimization.

Using initial estimates $\tilde{\beta}$, adaptive weights are calculated as follows:

$$\tilde{w}_j = \begin{cases} \frac{1}{\tilde{\beta}_j} & \text{if } \tilde{\beta}_j \neq 0 \\ \infty & \text{if } \tilde{\beta}_j = 0 \end{cases} \tag{4}$$

(with the convention $0 \cdot \infty = 0$).

The Adaptive Lasso Refinement is as follows:

$$L_2(\theta; r) = \ln L(\lambda_0, \beta, \sigma^2) + r \sum_{j=1}^{p} \tilde{w}_j |\beta_j|. \tag{5}$$

This two-step approach enhances variable selection consistency while mitigating the over-penalization of significant coefficients.

2.3. Differential Privacy

We start by introducing the concept of differential privacy.

Definition 1 (Differential Privacy [11]). *A randomized mechanism $M : \mathcal{X} \to \mathcal{Y}$ is said to be (ϵ, δ)-differentially private for positive parameters ϵ, δ. If, for any pair of adjacent datasets $X, X' \in \mathcal{X}$ differs by a single record, and for every measurable subset $\mathcal{T} \subseteq \mathcal{Y}$ under $M(\cdot)$, the following holds:*

$$\mathbb{P}[M(X) \in \mathcal{T}] \leq e^{\epsilon} \mathbb{P}[M(X') \in \mathcal{T}] + \delta, \tag{6}$$

where the probability $\mathbb{P}$ arises solely from the randomness of the mechanism M.

In this definition, the datasets X and X' are considered fixed, with the probability capturing the stochastic nature of $M(\cdot)$. The privacy level against potential adversaries is determined by the parameters ϵ and δ. Smaller values of these parameters indicate stronger privacy protections.

A fundamental principle in differential privacy is sensitivity, which measures the maximum impact on an algorithm's output resulting from the modification of a single record in the dataset.

Definition 2 (Sensitivity). *For a deterministic, vector-valued function $\mathcal{F}(\cdot) : X \in \mathcal{X} \to \mathbb{R}^n$, the l_p sensitivity of $\mathcal{F}(\cdot)$ is given as follows:*

$$\Delta_p(\mathcal{F}) := \sup_{X, X' \in \mathcal{X}} \|\mathcal{F}(X) - \mathcal{F}(X')\|_p, \tag{7}$$

where X and X' differ by only one entry.

We proceed to describe the Laplace mechanism, which is a well-known technique that guarantees (ϵ, δ)-differential privacy for functions exhibiting constrained l_1 sensitivity.

Lemma 1 (Laplace Mechanism [31]). *For a function $g : \mathcal{X} \to \mathbb{R}^n$ with l_1-sensitivity $\Delta_1(g)$, the mechanism $M(X) = g(X) + Lap(\sigma)$, where $Lap(\sigma) \sim Laplace(0, \sigma)$, ensures ϵ-differential privacy, provided $\sigma = \Delta_1(g)/\epsilon$.*

Lemma 2 (Advanced Composition [32]). *For a sequence of T mechanisms, each satisfying ϵ'-differential privacy, their composition satisfies (ϵ, δ)-differential privacy, where the following is true:*

$$\epsilon = \epsilon' \sqrt{2T \ln(1/\delta)} + T\epsilon'(e^{\epsilon'} - 1). \tag{8}$$

2.4. Constrained Parameter Space and Sensitivity Analysis

Our first methodological innovation is the formulation of a convex constrained parameter space for the SFM. This design plays a dual role—(i) it guarantees numerical stability during optimization by preventing extreme parameter realizations, and (ii) it ensures bounded gradient sensitivity, which is indispensable for the enforcement of differential privacy in a non-convex setting.

We define the optimization objective via the negative log-likelihood, as follows:

$$\mathcal{L}(\theta) = -\frac{1}{N} \ln L(\lambda_0, \beta, \sigma^2), \tag{9}$$

where $\theta = \{\lambda_0, \beta, \sigma^2\}$. By introducing parameter constraints, the feasible set is restricted to a convex region, thereby avoiding unstable or ill-posed solutions.

Definition 3 (Constrained Parameter Space). *The feasible parameter set is defined as a convex polytope $\mathcal{C} \subset \mathbb{R}^{p+2}$, as follows:*

$$\mathcal{C} = \left\{ (\beta, \sigma^2, \lambda_0) \in \mathbb{R}^{p+2} \,\middle|\, \begin{array}{l} \|\beta\|_1 \le B, \\ \sigma^2 \in [\sigma^2_{\min}, \sigma^2_{\max}], \\ \lambda_0 \in [0, \lambda_{\max}] \end{array} \right\},$$

where $B, \sigma^2_{\min}, \sigma^2_{\max}, and \lambda_{\max} > 0$ are fixed constants.

These bounds simultaneously guarantee numerical stability and finite sensitivity. Unlike convex optimization problems, where the Lipschitz continuity of the gradient is sufficient for privacy guarantees, the non-convexity induced by the $\Phi(\cdot)$ term in SFM complicates such analysis. Our approach circumvents this difficulty by focusing directly on bounding the sensitivity of the gradient.

Gradient Sensitivity Analysis

The sensitivity of the gradient is highly dependent on the bounds of λ_0 and σ^2. From the derivations provided in Appendix A, the partial derivative with respect to λ_0 exhibits the following scaling:

$$\mathcal{O}\left(\frac{\lambda_0}{\sigma_{\min}} \cdot M(Z_{\max}) \right), \tag{10}$$

where $Z_{\max} = \frac{(1+B)\lambda_0}{\sigma_{\min}}$ and $M(Z_{\max}) = \max\left\{ \frac{2}{\sqrt{2\pi}}, 2Z_{\max} + 2 \right\}$. This behavior implies that the gradient grows nearly linearly when λ_0 is small, but increases superlinearly as λ_0 becomes larger, primarily due to the Mills ratio.

Similarly, the partial derivative with respect to σ^2 satisfies the following:

$$\mathcal{O}\left(\sigma^{-2} + \sigma^{-4} \right), \tag{11}$$

indicating that excessively small values of σ^2 cause gradient explosion. This effect not only destabilizes optimization but also amplifies the noise required by differentially private mechanisms.

Taken together, these analyses reveal a sharp trade-off, whereby smaller σ^2 and larger λ_0 improve model flexibility but substantially increase gradient sensitivity. Our convex constrained parameter design directly addresses this issue by imposing explicit bounds on both parameters. This constitutes the first methodological contribution of the paper, as it establishes a feasible optimization domain that ensures bounded sensitivity—a prerequisite for developing a differentially private stochastic frontier estimator.

2.5. Private SFM

This subsection presents our primary algorithmic contribution—a differentially private estimation framework for SFM, based on the Frank–Wolfe procedure within the previously defined constrained parameter space. The method leverages convex–analytic techniques to balance statistical efficiency with privacy guarantees.

The Frank–Wolfe algorithm [33] is a first-order method for constrained convex optimization. At each iteration, the algorithm linearizes the objective function around the current iterate and then updates the solution by moving toward the minimizer of the linearized problem over the convex domain. Formally, given the empirical loss function $\mathcal{L}$ defined in (9), the optimization problem can be written as follows:

$$\min_{\theta \in \mathcal{C}} \ \mathcal{L}(\theta). \tag{12}$$

Due to the presence of the inefficiency term u_i, the objective $\mathcal{L}(\theta)$ is generally non-convex, as the cumulative distribution term $\Phi(\cdot)$ introduces nonlinearities [2]. Nonetheless, the Frank–Wolfe procedure (Algorithm 1) remains a suitable vehicle for our private adaptation.

Algorithm 1 Frank–Wolfe Procedure

Input: Feasible domain $\mathcal{C} \subseteq \mathbb{R}^{p+2}$, objective function $\mathcal{L} : \mathcal{C} \to \mathbb{R}$, learning rate schedule $\{\mu_t\}$

1: Initialize $\theta^{(1)} \in \mathcal{C}$
2: **for** $t = 1$ to $T - 1$ **do**
3: Solve linear subproblem: $s^{(t)} = \arg\min_{\theta \in \mathcal{C}} \langle \nabla \mathcal{L}(\theta^{(t)}), \theta \rangle$
4: Update iterate: $\theta^{(t+1)} = (1 - \mu_t)\theta^{(t)} + \mu_t s^{(t)}$
5: **return** Final output $\theta^{(T)}$

The convergence rate of Frank–Wolfe depends on the curvature of $\mathcal{L}$. Intuitively, this curvature quantifies how much $\mathcal{L}$ deviates from its linear approximation over the domain $\mathcal{C}$.

Definition 4 (Curvature Constant)**.** *The curvature constant of $\mathcal{L}$ over $\mathcal{C}$ is defined as follows:*

$$\Gamma_{\mathcal{L}} = \sup_{\theta_1, \theta_2 \in \mathcal{C}, \gamma \in (0,1), \theta_3 = \theta_1 + \gamma(\theta_2 - \theta_1)} \frac{2}{\gamma^2} [\mathcal{L}(\theta_3) - \mathcal{L}(\theta_1) - \langle \nabla \mathcal{L}(\theta_1), \theta_3 - \theta_1 \rangle]. \tag{13}$$

To safeguard individual-level data during SFM estimation, we embed differential privacy into the Frank–Wolfe framework by adapting the differentially private Frank–Wolfe (DP-FW) algorithm proposed by Bassily et al. [34]. In each iteration t, the gradient direction is selected using the exponential mechanism, with Laplace noise being calibrated to the ℓ_1 sensitivity of the gradient. This ensures that private information about any single observation is obfuscated.

Although $\mathcal{L}(\theta)$ is non-convex, DP-FW remains an effective heuristic, since privacy guarantees do not rely on convexity. Furthermore, the restricted parameter space $\mathcal{C}$, being convex and polyhedral, enables efficient vertex-based updates.

The privacy budget depends critically on bounding the sensitivity of the gradient $\nabla\mathcal{L}(\theta; D)$, where $D = \{(x_i, y_i)\}_{i=1}^N$ denotes the dataset.

Theorem 1 (Gradient Sensitivity)**.** *The ℓ_1 sensitivity of $\nabla\mathcal{L}(\theta; D)$ is as follows:*

$$\Delta = \max_{\theta\in\mathcal{C}} \max_{D,D'} \|\nabla\mathcal{L}(\theta; D) - \nabla\mathcal{L}(\theta; D')\|_1 \leq \frac{2G}{N}, \tag{14}$$

where $G = \sup_{\theta\in\mathcal{C},d} \|\nabla\ell(\theta; d)\|_1 < \infty$.

Proof. Let D and D' be adjacent datasets differing in one record. Since the following is true:

$$\nabla\mathcal{L}(\theta; D) = \frac{1}{N}\sum_{i=1}^{N} \nabla\ell(\theta; d_i),$$

the difference between gradients under D and D' reduces to the following:

$$\nabla\mathcal{L}(\theta; D) - \nabla\mathcal{L}(\theta; D') = \frac{1}{N}\big(\nabla\ell(\theta; d_k) - \nabla\ell(\theta; d'_k)\big).$$

Thus,

$$\|\nabla\mathcal{L}(\theta; D) - \nabla\mathcal{L}(\theta; D')\|_1 \leq \frac{1}{N}\big(\|\nabla\ell(\theta; d_k)\|_1 + \|\nabla\ell(\theta; d'_k)\|_1\big) \leq \frac{2}{N}\sup_{\theta\in\mathcal{C},d} \|\nabla\ell(\theta; d)\|_1.$$

With bounded covariates ($\|x_i\|_\infty \leq 1$), bounded outputs ($|y_i| \leq 1$), and constraints on $\mathcal{C}$, a finite constant G exists such that $\sup_{\theta,d} \|\nabla\ell(\theta; d)\|_1 \leq G$. Hence, $\Delta \leq 2G/N$. □

Incorporating the sensitivity bound, we adapt the DP-FW algorithm to the SFM setting (Algorithm 2). At each iteration, noisy gradient scores are computed for all candidate vertices using the exponential mechanism, and the update is taken toward the vertex with the lowest perturbed score.

Algorithm 2 Differentially Private Frank–Wolfe for SFM.

Input: Data $D = \{(x_i, y_i)\}_{i=1}^N$, loss $\mathcal{L}(\theta; D)$, privacy budget (ϵ, δ), feasible domain $\mathcal{C} = \text{conv}(S)$, steps T

Output: Private estimate θ_{priv}

1: Initialize $\theta^{(1)} \in \mathcal{C}$
2: **for** $t = 1$ to $T - 1$ **do**
3: **for all** $s \in S$ **do**
4: Compute noisy score:

$$\alpha_s^{(t)} \leftarrow \langle s, \nabla\mathcal{L}(\theta^{(t)}; D)\rangle + \text{Lap}\left(\frac{2G \cdot \|\mathcal{C}\|_1 \cdot \sqrt{8T\log(1/\delta)}}{N\epsilon}\right)$$

5: Select vertex: $s^{(t)} = \arg\min_{s\in S} \alpha_s^{(t)}$
6: Update: $\theta^{(t+1)} = (1 - \mu_t)\theta^{(t)} + \mu_t s^{(t)}$, with $\mu_t = \frac{2}{t+2}$
7: **return** $\theta_{\text{priv}} = \theta^{(T)}$

Here, $\|\mathcal{C}\|_1 = \max_{s\in\mathcal{S}} \|s\|_1$, and the noise calibration ensures (ϵ, δ)-differential privacy. This integration of Frank–Wolfe with differential privacy constitutes the second methodological contribution of this work.

2.6. Privacy Guarantee

This subsection establishes the formal privacy properties of our proposed algorithm, proving that it satisfies (ϵ, δ)-differential privacy. This constitutes a central theoretical contribution of the paper.

Theorem 2 (Laplace Noise Scale)**.** *In Algorithm 2, the required Laplace noise scale λ is given as follows:*

$$\lambda = \frac{2G\|\mathcal{C}\|_1\sqrt{8T\log(1/\delta)}}{N\epsilon}, \tag{15}$$

where $\|\mathcal{C}\|_1 = \max_{s\in\mathcal{S}} \|s\|_1$.

Proof. Consider the score function $\alpha_s = \langle s, \nabla\mathcal{L}(\theta; D)\rangle$. Its sensitivity is as follows:

$$\Delta_\alpha = \max_{s\in S}\max_{D,D'} |\langle s, \nabla\mathcal{L}(\theta; D) - \nabla\mathcal{L}(\theta; D')\rangle| \leq \max_{s\in S} \|s\|_1 \cdot \frac{2G}{N} = \frac{2G\|\mathcal{C}\|_1}{N}.$$

According to the Laplace mechanism (Lemma 1), ensuring ϵ'-differential privacy per iteration requires the following:

$$\lambda = \frac{\Delta_\alpha}{\epsilon'} = \frac{2G\|\mathcal{C}\|_1}{N\epsilon'}.$$

To guarantee an overall (ϵ, δ)-differential privacy budget over T iterations, we invoke the advanced composition theorem (Lemma 2), setting the following:

$$\epsilon' = \frac{\epsilon}{\sqrt{8T\log(1/\delta)}}.$$

Substituting yields the claimed result. □

We now establish the main theoretical guarantee for Algorithm 2.

Theorem 3. *Algorithm 2 satisfies (ϵ, δ)-differential privacy.*

Proof. The proof builds on the analysis of the private Frank–Wolfe algorithm proposed by Talwar et al. [15]. At each iteration t, the algorithm selects a direction $\tilde{\theta}_t$ using the exponential mechanism with the following score function:

$$\alpha_s = \langle s, \nabla\mathcal{L}(\theta_t; D)\rangle + \mathrm{Lap}(\lambda),$$

where $\lambda = \frac{2G\|\mathcal{C}\|_1\sqrt{8T\log(1/\delta)}}{N\epsilon}$.

We first verify per-iteration privacy. For adjacent datasets D and D', the sensitivity of the score function is as follows:

$$\max_{D,D'} |\langle s, \nabla\mathcal{L}(\theta_t; D) - \nabla\mathcal{L}(\theta_t; D')\rangle|.$$

According to the Cauchy–Schwarz inequality in the ℓ_1-norm, the following is true:

$$|\langle s, \nabla\mathcal{L}(\theta_t; D) - \nabla\mathcal{L}(\theta_t; D')\rangle| \leq \|s\|_1 \cdot \|\nabla\mathcal{L}(\theta_t; D) - \nabla\mathcal{L}(\theta_t; D')\|_1.$$

From Theorem 1, the following holds true:

$$\|\nabla\mathcal{L}(\theta_t; D) - \nabla\mathcal{L}(\theta_t; D')\|_1 \leq \frac{2G}{N},$$

and since $s \in \mathcal{C}$, we have $\|s\|_1 \leq \|\mathcal{C}\|_1$. Thus, the overall sensitivity is bounded as follows:

$$\max_{D,D'} |\langle s, \nabla\mathcal{L}(\theta_t; D) - \nabla\mathcal{L}(\theta_t; D')\rangle| \leq \frac{2G\|\mathcal{C}\|_1}{N}.$$

To ensure ϵ'-differential privacy per iteration, the Laplace noise must satisfy the following:

$$\lambda \geq \frac{\text{sensitivity}}{\epsilon'} = \frac{2G\|\mathcal{C}\|_1}{N\epsilon'}.$$

Setting $\epsilon' = \frac{\epsilon}{\sqrt{8T\log(1/\delta)}}$ gives the following:

$$\lambda \geq \frac{2G\|\mathcal{C}\|_1\sqrt{8T\log(1/\delta)}}{N\epsilon},$$

matching the chosen λ in Algorithm 2.

Finally, applying the advanced composition theorem, the cumulative privacy loss after T iterations is as follows:

$$\epsilon_{\text{total}} = \sqrt{2T\ln(1/\delta)}\cdot\epsilon' + T\cdot\epsilon'(e^{\epsilon'} - 1).$$

Substituting $\epsilon' = \frac{\epsilon}{\sqrt{8T\log(1/\delta)}}$, we obtain the following:

$$\sqrt{2T\ln(1/\delta)}\cdot\epsilon' = \frac{\epsilon}{2}.$$

For small ϵ', $e^{\epsilon'} - 1 \approx \epsilon'$, implying the following:

$$T\cdot\epsilon'(e^{\epsilon'} - 1) \approx \frac{\epsilon^2}{8\log(1/\delta)}.$$

Hence,

$$\epsilon_{\text{total}} \approx \frac{\epsilon}{2} + \frac{\epsilon^2}{8\log(1/\delta)} \leq \epsilon, \tag{16}$$

when δ is sufficiently small and ϵ is moderate. Therefore, Algorithm 2 satisfies (ϵ, δ)-differential privacy. Importantly, this guarantee holds irrespective of the potential non-convexity of $\mathcal{L}(\theta)$, as it relies solely on bounded sensitivity and the Laplace mechanism. □

3. Numerical Experiments

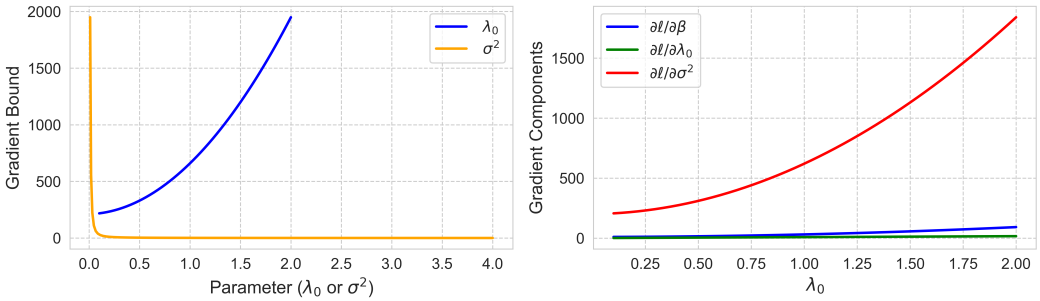
This section evaluates the proposed DP-FW framework through sensitivity analysis, simulation studies, and real-world data applications, contributing original empirical evidence to validate its effectiveness and robustness in SFA.

3.1. Gradient Sensitivity Analysis

To validate our theoretical gradient bounds and demonstrate the practical impact of parameter constraints, we conduct a comprehensive sensitivity analysis examining how λ_0 and σ^2 affect gradient magnitude and numerical stability.

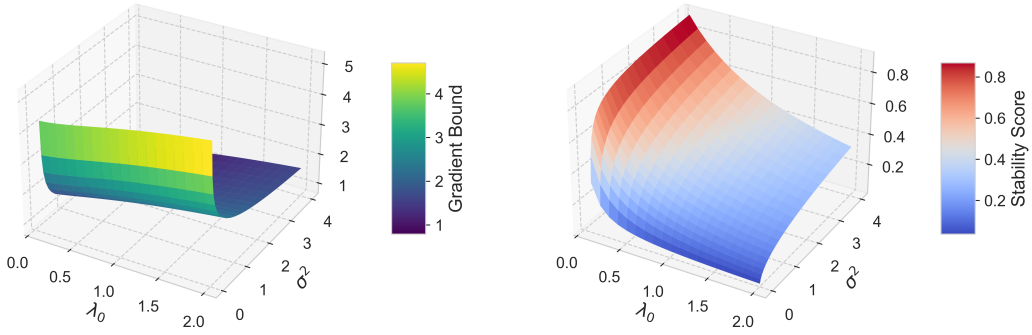
Specifically, we fix $B = 1$ and $p = 5$, while varying $\lambda_0 \in [0.1, 5.0]$ and $\sigma^2 \in [0.01, 2.0]$. We first examine the influence curves of λ_0 and σ^2 on gradient components (Figure 1),

and then provide a spatial heatmap visualization of sensitivity and stability across the parameter space (Figure 2; see Appendix Stability Metrics for stability score definitions).



(a) Gradient Sensitivity to λ_0 and σ^2 (b) Relative Contribution of Gradient Components

Figure 1. Gradient-bound sensitivity with respect to λ_0 and σ^2. (**a**) Overall gradient bound vs. parameter values. (**b**) Gradient components with respect to β, λ_0, and σ^2.



(a) Gradient Sensitivity (b) Numerical Stability

Figure 2. Parameter space visualization of gradient sensitivity and numerical stability. (**a**) Gradient-bound surface. (**b**) Stability score surface across (λ_0, σ^2).

Figure 1 illustrates that the gradient component associated with λ_0 grows exponentially when $\lambda_0 > 0.45$, confirming the Mills ratio effect. For σ^2, explosive sensitivity occurs near zero, with the gradient magnitude increasing by up to $4800\times$ when $\sigma^2 < 0.1$. Overall, σ^2 dominates the gradient dynamics across most practical regimes.

The heatmaps in Figure 2 further highlight the interplay between λ_0 and σ^2, visually confirming regions of instability. These results empirically validate our theoretical framework and provide practical guidance for parameter selection in differentially private implementations.

3.2. Simulation Study

We next investigate the statistical inference and variable selection capabilities of DP-FW under controlled simulation settings.

Our study addresses two fundamental questions—(1) How effectively does the proposed method estimate parameters across different sample sizes and privacy constraints? (2) Can the framework maintain variable selection accuracy in high-dimensional sparse settings while preserving privacy?

We benchmark DP-FW against maximum likelihood estimation (MLE), classical Frank–Wolfe (FW), stochastic gradient descent (SGD), and differentially private variants (DP-SGD and DP-Bayesian). The initial estimation focuses on DP-FW and DP-SGD to contrast Frank–Wolfe with gradient-based optimization, while DP-Bayesian is introduced later due to its computational cost and greater relevance in small-sample or high-dimensional settings.

Evaluation metrics include Mean Squared Error (MSE), True Positive Rate (TPR), False Positive Rate (FPR), and Sum of Squared Errors (SSE).

3.2.1. Parameter Estimation

We adopt the following linear stochastic frontier model:

$$Y = X\beta + \epsilon, \quad \epsilon = v - u, \tag{17}$$

where $v \sim N(0, \sigma_v^2 I_N)$ represents symmetric noise, and $u \sim N^+(0, \sigma_u^2 I_N)$ denotes non-negative inefficiency, both independent across N observations.

In low-dimensional settings, we fix $s = 3$ non-zero coefficients and $p = 3$, with variance parameters $\sigma_v^2 = 0.25$ and $\sigma_u^2 = 0.09$, as well as true coefficients $\beta = (1.0, 2.0, 3.0)^\top$. The sample size N is varied from 500 to 5000 in increments of 500, while privacy parameters are set to $\epsilon = 1$ and $\delta = 10^{-5}$. Hyperparameters are tuned via 5-fold cross-validation, with MSE as the criterion, using Optuna. Each experiment is repeated 100 times to mitigate randomness. For brevity, Table 1 reports representative results for $N \in \{500, 1500, 3000\}$.

Table 1. Comparison of parameter estimates and standard deviations across methods ($\epsilon = 1$).

N	Param. (True)	Method									
		MLE		**FW**		**DP-FW**		**SGD**		**DP-SGD**	
		Est	**SD**	**Est**	**SD**	**Est**	**SD**	**Est**	**SD**	**Est**	**SD**
500	β_1 (1.0)	0.980	**0.057**	0.985	0.061	0.985	0.060	**1.006**	0.063	0.948	0.072
	β_2 (2.0)	**2.011**	**0.083**	2.089	0.091	2.064	0.087	1.910	0.092	2.039	0.105
	β_3 (3.0)	**3.016**	**0.104**	3.148	0.113	3.113	0.108	3.120	0.115	2.980	0.130
	σ^2 (0.34)	**0.334**	**0.033**	0.381	0.038	0.449	0.041	0.330	0.034	0.460	0.045
	λ_0 (0.6)	0.425	**0.098**	0.732	0.105	**0.681**	0.102	0.430	0.100	0.750	0.110
	MSE	**0.326**		0.409		0.399		0.460		0.570	
1500	β_1 (1.0)	1.078	**0.032**	**1.012**	0.035	1.025	0.034	0.916	0.035	1.015	0.041
	β_2 (2.0)	2.013	**0.047**	2.044	0.051	2.043	0.050	2.048	0.052	**1.988**	0.060
	β_3 (3.0)	**2.997**	**0.064**	2.968	0.067	2.963	0.066	2.975	0.070	2.992	0.079
	σ^2 (0.34)	0.351	**0.020**	0.411	0.021	0.420	0.023	**0.348**	**0.020**	0.430	0.025
	λ_0 (0.6)	0.746	**0.063**	1.002	0.072	1.069	0.069	**0.740**	0.065	1.080	0.075
	MSE	0.347		0.348		0.350		**0.310**		0.348	
3000	β_1 (1.0)	0.970	**0.022**	1.049	0.024	1.016	0.024	**1.012**	0.024	1.067	0.028
	β_2 (2.0)	**1.993**	**0.033**	1.959	0.035	1.965	0.035	1.992	0.036	2.015	0.042
	β_3 (3.0)	**3.000**	**0.042**	2.933	0.045	2.944	0.043	2.991	0.046	2.994	0.052
	σ^2 (0.34)	0.326	**0.013**	0.403	0.015	0.367	0.014	**0.327**	0.014	0.380	0.016
	λ_0 (0.6)	0.576	**0.042**	0.915	0.048	0.819	0.046	**0.580**	**0.042**	0.850	0.050
	MSE	**0.319**		0.344		0.338		0.368		0.418	

Table 1 shows that while all methods accurately estimate the structural coefficients (β), MLE consistently achieves the lowest bias and variance, especially for the error parameters (σ^2 and λ_0). Among the private methods, our proposed framework not only estimates β almost as well as the MLE baseline, but also decisively outperforms unconstrained private gradient methods in overall accuracy.

Sample Size and Privacy Effects

We further analyze sensitivity to data availability and privacy strength. Figure 3 reports MSE trends as N increases under $\epsilon = 1$. Figure 4 explores the privacy–utility trade-off by varying ϵ from 0.1 to 1 at fixed $N = 5000$.

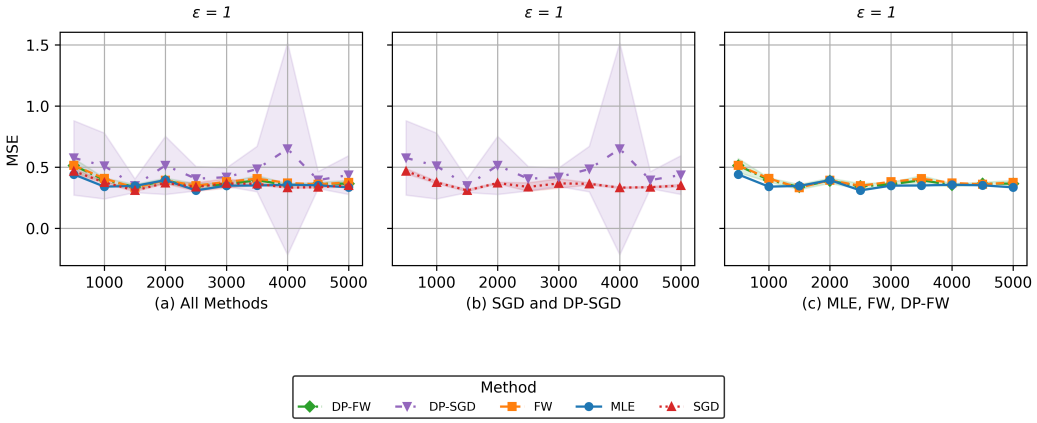


Figure 3. Impact of data size on MSE at $\epsilon = 1$. (**a**) All methods, including private and non-private ones. (**b**) Stochastic gradient-based methods: SGD and DP-SGD. (**c**) Stable methods.

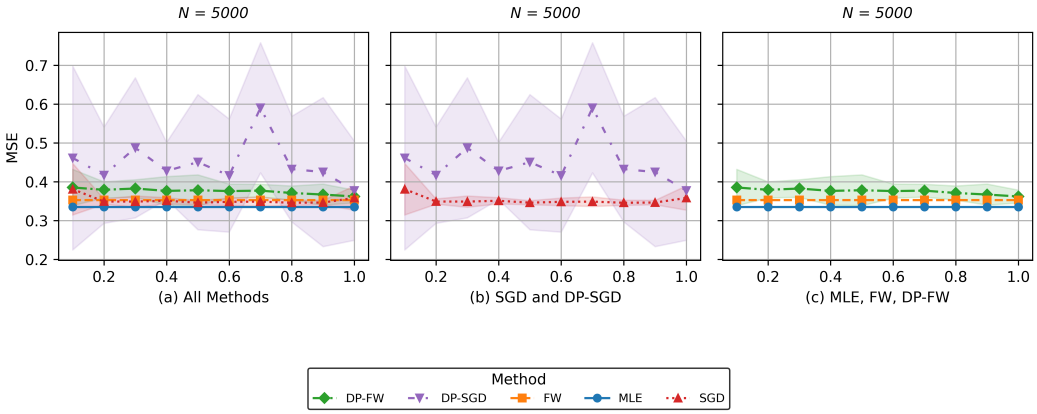


Figure 4. Privacy–utility trade-off with varying ϵ at $N = 5000$. (**a**) All methods, including private and non-private ones. (**b**) Stochastic gradient-based methods: SGD and DP-SGD. (**c**) Stable methods.

Figure 3 shows that increasing N consistently reduces MSE across all methods. While non-private estimators converge stably, private gradient-based methods exhibit volatility at moderate sample sizes. In contrast, DP-FW scales smoothly with N, confirming that bounded optimization effectively suppresses gradient sensitivity and facilitates the efficient use of additional data.

Figure 4 demonstrates the privacy–utility trade-off. As ϵ increases, all private methods benefit from reduced noise, but DP-FW shows a markedly smoother improvement curve. This stability arises from the controlled gradient domain, which ensures that injected perturbations scale predictably with privacy level. The result is better utility per unit of privacy budget—a critical advantage in practical applications.

Small-Sample Robustness

Small-sample scenarios are common in specialized efficiency analysis applications where data availability is inherently limited, posing a greater challenge for differentially private methods due to the amplified effect of injected noise.

To examine robustness, we conduct experiments with $N \in \{50, 100, 150\}$ under $\epsilon = 1$, including DP-Bayesian methods for comparison. The results are summarized in Figure 5 and Table 2.

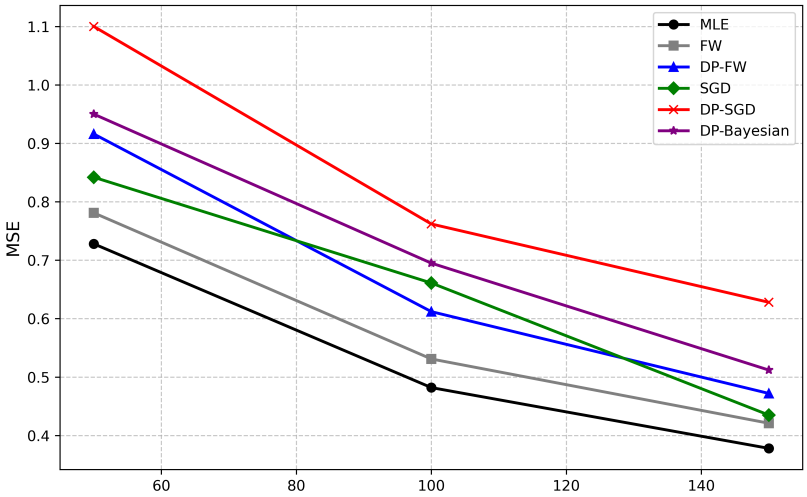


Figure 5. Comparative performance of different methods in small-sample regimes for $\epsilon = 1$.

Table 2. Parameter estimates and standard deviations in small-sample regimes ($\epsilon = 1$).

N	Param. (True)	Method											
		MLE		FW		SGD		DP-FW		DP-SGD		DP-Bayesian	
		Est	SD	Est	SD	Est	SD	Est	SD	Est	SD	Est	SD
50	β_1 (1.0)	0.952	**0.185**	1.089	0.203	1.124	0.218	1.078	0.196	0.847	0.245	**1.015**	0.208
	β_2 (2.0)	**2.034**	**0.268**	2.156	0.295	2.245	0.318	2.142	0.285	1.785	0.352	2.089	0.301
	β_3 (3.0)	**3.068**	**0.342**	3.234	0.378	3.185	0.395	3.156	0.368	2.892	0.445	3.201	0.389
	σ^2 (0.34)	**0.348**	**0.108**	0.425	0.125	0.398	0.132	0.465	0.118	0.512	0.148	0.478	0.128
	λ_0 (0.6)	0.385	**0.285**	0.892	0.325	0.425	0.318	**0.735**	0.295	0.965	0.385	0.825	0.342
	MSE	**0.728**		0.781		0.842		0.916		1.100		0.950	
100	β_1 (1.0)	**0.976**	**0.128**	1.045	0.142	1.089	0.156	1.032	0.138	0.915	0.175	1.068	0.148
	β_2 (2.0)	**2.018**	**0.185**	2.098	0.205	2.134	0.225	2.076	0.198	1.892	0.248	2.125	0.215
	β_3 (3.0)	**3.042**	**0.238**	3.156	0.265	3.098	0.285	3.089	0.258	2.945	0.318	3.142	0.275
	σ^2 (0.34)	**0.342**	**0.075**	0.398	0.088	0.385	0.095	0.425	0.085	0.485	0.108	0.445	0.092
	λ_0 (0.6)	0.425	**0.198**	0.785	0.225	0.458	0.218	**0.695**	0.205	0.825	0.268	0.758	0.238
	MSE	**0.482**		0.531		0.661		0.612		0.762		0.695	
150	β_1 (1.0)	**0.985**	**0.105**	1.028	0.118	1.056	0.128	1.015	0.115	0.948	0.142	1.042	0.122
	β_2 (2.0)	**2.012**	**0.152**	2.068	0.168	2.085	0.185	2.045	0.165	1.925	0.205	2.078	0.175
	β_3 (3.0)	**3.025**	**0.195**	3.098	0.218	3.065	0.235	3.058	0.212	2.985	0.258	3.089	0.225
	σ^2 (0.34)	**0.338**	**0.062**	0.385	0.072	0.375	0.078	0.405	0.070	0.455	0.088	0.425	0.075
	λ_0 (0.6)	0.485	**0.162**	0.725	0.185	0.512	0.178	**0.668**	0.168	0.785	0.215	0.715	0.192
	MSE	**0.378**		0.421		0.435		0.472		0.628		0.512	

Table 2 and Figure 5 reveal several key insights. As anticipated, all methods show improved accuracy with increasing sample size. However, our proposed framework demonstrates superior robustness. Notably, at $N = 100$, DP-FW achieves a good performance for structural parameters, which is comparable to the non-private SGD method.

This counterintuitive result—where private methods occasionally outperform non-private alternatives—arises from two complementary mechanisms. First, gradient descent

methods exhibit a higher variance in small-sample settings, potentially leading to overfitting or unstable convergence paths. Second, bounded-domain optimization combined with calibrated perturbation provides implicit regularization effects that can improve generalization performance when data are limited.

Privacy–Utility Trade-Off

Building upon the previous analysis of sample size effects under a fixed privacy budget, we now examine how model performance changes with varying privacy parameters ϵ.

Figure 6 reports the MSE of differentially private estimators as the privacy budget ϵ varies. Unconstrained DP-SGD exhibits inflated MSE and instability under tight privacy constraints, reflecting the difficulty of controlling gradient sensitivity in high-dimensional settings. In contrast, both Bayesian approaches and DP-FW remain substantially more stable across privacy levels. Importantly, DP-FW consistently achieves a lower MSE than Bayesian alternatives, as well as displaying smoother improvements as ϵ increases.

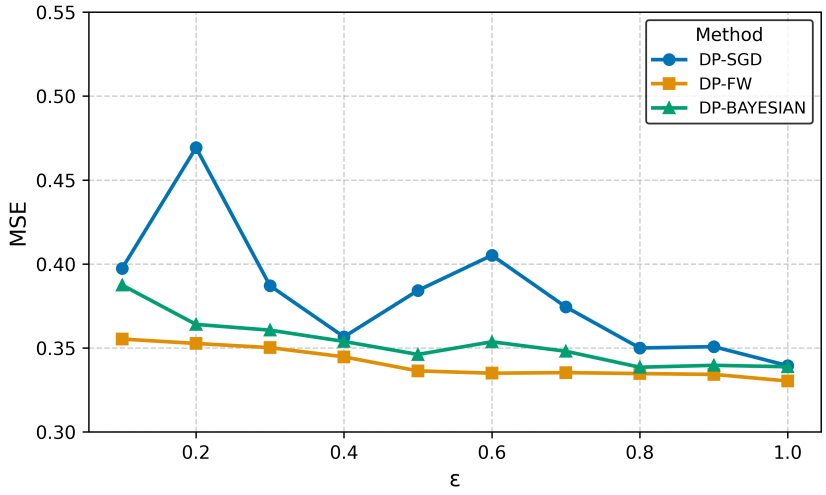


Figure 6. Privacy–utility trade-off with varying ϵ at $N = 3000$. Comparison among DP methods.

Taken together, these results highlight how bounding the optimization domain translates into predictable sensitivity and improved variance control. Although differentially private estimators inevitably lose some efficiency relative to their non-private counterparts due to noise injection, DP-FW mitigates much of this loss by converting perturbation into a form of implicit regularization. The benefits are most pronounced in challenging regimes—small sample sizes, high-dimensional covariates, and stringent privacy budgets—where domain constraints and calibrated noise jointly stabilize estimation, narrow the gap to non-private baselines, and consistently outperform alternative private methods.

3.2.2. Variable Selection

This subsection evaluates the variable selection performance of our proposed DP-FW-Lasso in high-dimensional sparse settings, providing original evidence of its effectiveness under privacy constraints.

We consider a design with $p = 500$ predictors, of which only $s = 6$ coefficients are non-zero; these are randomly set to ± 1. To promote sparsity, we incorporate LASSO regularization into both the non-private Frank–Wolfe estimator and its private counterpart (DP-FW-Lasso). This integration is natural, as Frank–Wolfe directly accommodates ℓ_1-type constraints. As a non-private benchmark, we include Adaptive LASSO (ALasso) with MLE, which is well known for its variable selection consistency. We deliberately exclude a private version of ALasso, as its reliance on accurate initial estimates and iterative reweighting amplifies the effects of privacy-induced noise, reducing stability and complicating privacy

budget allocation. In contrast, standard LASSO is more amenable to privacy-preserving optimization, ensuring both computational tractability and stability.

Hyperparameters were tuned using the Bayesian Information Criterion (BIC) and 5-fold cross-validation. Performance was assessed using TPR, FPR, SSE, and MSE. Table 3 reports the results for $N = 5000$ and $\epsilon = 1$, with graphical comparisons provided in Figure 7.

Table 3. Variable selection performance metrics.

Metric	MLE-ALasso	FW-Lasso	DP-FW-Lasso	SGD-Lasso	DP-SGD-Lasso	DP-Bayesian
TPR	1.00	1.00	1.00	1.00	1.00	1.00
FPR	0.22	0.17	0.52 *	0.49	0.73	0.56
SSE	0.02	0.21	0.47 *	0.34	0.51	0.50
MSE	0.30	0.36	0.49 *	0.33	0.54	0.52

Note: within DP methods, values marked with * indicate the best performance.

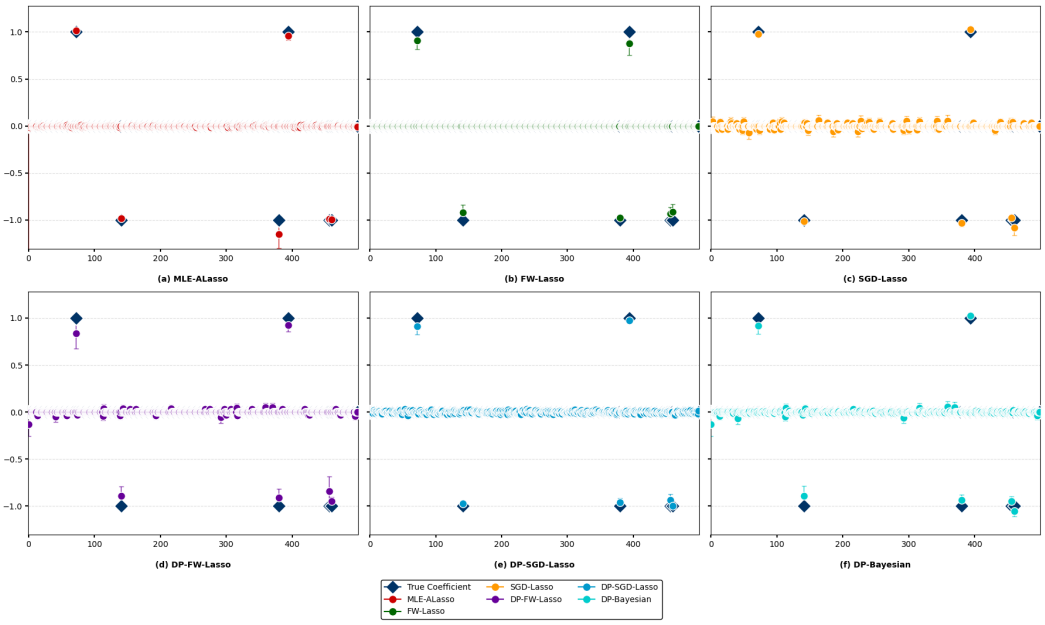


Figure 7. Variable selection performance.

Table 3 shows that all methods recover the true variables (TPR = 1.00) but differ in false positive control and estimation accuracy. Among non-private estimators, FW-LASSO produces the most parsimonious models. Under privacy constraints, DP-FW-Lasso attains the lowest FPR and estimation error, outperforming DP-SGD-Lasso and DP-Bayesian while avoiding their scalability and stability issues.

Figure 7 further shows that DP-FW-Lasso better preserves sparsity and interpretability than other DP methods. These findings confirm the design objectives of our framework—by leveraging a convex constraint space, controlling gradient sensitivity, and integrating LASSO regularization, DP-FW-Lasso achieves a robust and superior balance between privacy, sparsity, and estimation accuracy in high-dimensional stochastic frontier models.

3.3. Real Data Analysis

We evaluate our differentially private stochastic frontier inference framework using two complementary datasets. The California Housing dataset serves as a methodological validation platform to assess algorithm performance in high-dimensional sparse regression under privacy constraints. In contrast, the Farm Accountancy Data Net-

work (FADN) dataset demonstrates a genuine application of SFA for measuring farm production efficiency.

The California Housing dataset (https://www.dcc.fc.up.pt/~ltorgo/Regression/cal_housing.html, accessed on 1 March 2025) comprises 20,640 observations with 8 features from the 1990 U.S. census. We expand these through nonlinear transformations, interactions, and spatial indicators to exceed 100 dimensions, creating a challenging setting for differentially private inference. The FADN dataset (http://ec.europa.eu/agriculture/rica/, accessed on 1 March 2025) tracks farm income and productivity across EU member states. Using a 2013 cross-section, we constructed a dataset with 51 features after handling missing values, outliers, and feature engineering. The model regresses total output on total input, agricultural land, and labor, with pronounced multicollinearity posing a rigorous test for inference methods.

We compare the following six estimators:

- MLE-ALasso: Non-private maximum likelihood estimation with adaptive LASSO.
- FW-Lasso: Non-private Frank–Wolfe estimator with ℓ_1 constraint.
- SGD-Lasso: Stochastic gradient descent with ℓ_1 penalty—non-private.
- DP-SGD-Lasso: Standard differentially private SGD with ℓ_1 penalty.
- DP-FW-Lasso: Proposed differentially private Frank–Wolfe estimator with ℓ_1 constraint.
- DP-Bayesian: Bayesian differentially private approach via posterior sampling.

All private methods ensure (ϵ, δ)-differential privacy with $\delta = 10^{-5}$. Each model is evaluated over five independent trials to assess variance, and prediction accuracy is measured via MAE and RMSE.

On the California Housing dataset, we assess methods under $\epsilon \in \{0.1, 0.5, 1.0\}$. Figure 8 shows that DP-FW-Lasso consistently outperforms other private methods, with its advantage growing under stricter privacy constraints where gradient-based approaches exhibit pronounced instability. Bayesian methods remain competitive at moderate privacy levels but are limited in high-dimensional scenarios due to scalability and computational demands. These results motivate the selection of $\epsilon = 0.1$ for subsequent real-world evaluation.

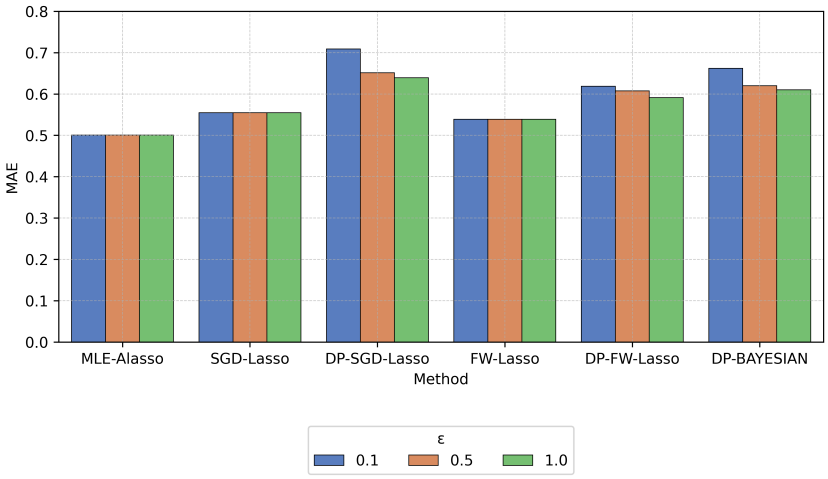


Figure 8. Prediction MAE under different privacy levels (ϵ) on the California Housing dataset.

Under the strictest privacy constraint ($\epsilon = 0.1$), we apply all methods to the FADN dataset. Figure 9 and Table 4 summarize the MAE and RMSE results. DP-FW-Lasso achieves a 35% lower MAE than DP-SGD-Lasso and a 12% improvement over DP-Bayesian. These gains reflect the superior handling of the privacy–utility trade-off, whereby Bayesian approaches are hindered by computational cost and limited scalability, while gradient descent suffers from optimization instability exacerbated by tight privacy constraints.

Table 4. Comparison of MAE and RMSE (Mean and Std) across different methods.

Method	MAE		RMSE	
	Mean	**Std**	**Mean**	**Std**
MLE-ALasso	0.500	0.001	0.688	0.001
FW-Lasso	0.539	0.013	0.695	0.010
SGD-Lasso	0.546	0.019	0.708	0.05
DP-SGD-Lasso	0.976	0.018	1.317	0.015
DP-Bayesian	0.720	0.001 *	0.882	0.001 *
DP-FW-Lasso	0.634 *	0.015	0.728 *	0.013

Note: Within DP methods, values marked with * indicate the best performance.

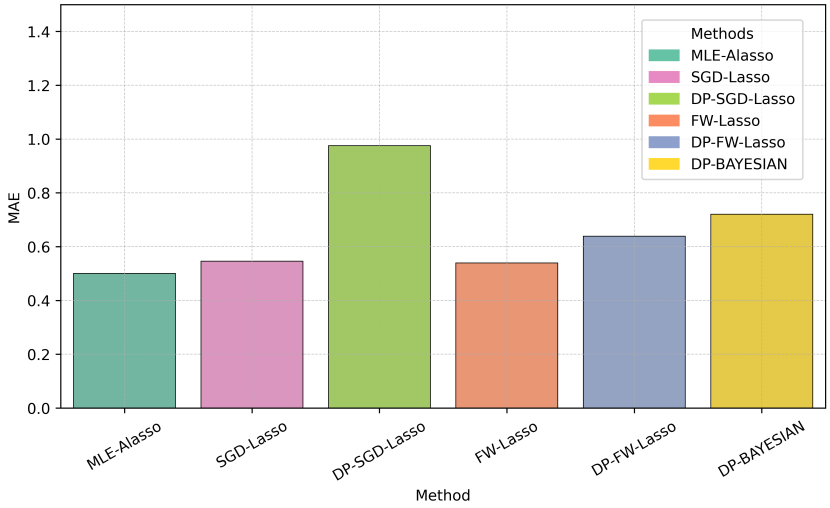


Figure 9. Prediction MAE under $\epsilon = 0.1$ on the FADN dataset across all methods.

Privacy-preserving methods inherently incur higher errors than non-private baselines, as noise injection for (ϵ, δ)-differential privacy increases estimation uncertainty in non-convex SFM. Frank–Wolfe's slightly higher errors relative to maximum likelihood reflect its focus on sparsity and variable selection over pure likelihood maximization, along with extra variance from mini-batch optimization.

Overall, across both datasets and privacy regimes, DP-FW-Lasso demonstrates consistent robustness, outperforming alternative private approaches in high-dimensional and structured econometric settings, while providing more scalable inference than computationally intensive Bayesian methods. These findings underscore the practical advantages of our framework for deploying differentially private stochastic frontier models in real-world applications.

4. Summary

This paper develops a novel framework for differentially private stochastic frontier analysis that enables reliable efficiency estimation under strict privacy protection. By bounding gradient sensitivity in non-convex frontier models and embedding these constraints into a Frank–Wolfe optimization scheme, we provide the first tractable approach for rigorous privacy-preserving efficiency analysis. The integration with ℓ_1 regularization further allows consistent variable selection in high-dimensional settings.

Empirical studies confirm the advantages of the proposed framework. Compared with existing private methods, it achieves 15–35% less prediction errors, maintains robustness under small samples and strong privacy constraints, and perfectly recovers sparsity in high-dimensional cases. Application to agricultural data further demonstrates that the method

preserves both accuracy and interpretability, making it suitable for practical economic efficiency evaluation under privacy restrictions.

While achieving substantial advances, our framework relies on convex relaxations and bounded parameter domains, which may limit applicability to highly flexible frontier specifications. The current cross-sectional focus also constrains application to dynamic efficiency tracking. Future research directions include extending the approach to panel data settings, developing adaptive noise calibration mechanisms for improved privacy–utility trade-offs, and exploring hybrid Bayesian–private inference for richer uncertainty quantification.

Author Contributions: Y.S.: study conception and design, development of methodology; M.Q.: data analysis, interpretation, and manuscript preparation and editing. X.W.: development of methodology. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Key Research and Development Program of China (2021YFA1000102), the Natural Science Foundation (NSF) project of Shandong Province of China (ZR2024MA074), the Ministry of education of Humanities and Social Science project (24YJA910003), and the Fundamental Research Funds for the Central Universities (No. 23CX03012A).

Data Availability Statement: This study has associated data that are deposited in data repositories.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Derivation of Gradient Bound G

This appendix provides a detailed derivation of the gradient bound G, which is essential for the sensitivity analysis in Theorem 1. The derivation relies on the following constraints on the parameter space and data:

- Bounded covariates: $\|x_i\|_\infty \leq 1$;
- Bounded outputs: $|y_i| \leq 1$;
- Constrained parameter space: $\|\beta\|_1 \leq B$, $\lambda_0 \in [0, \lambda_{\max}]$, $\sigma^2 \in [\sigma^2_{\min}, \sigma^2_{\max}]$ with $\sigma^2_{\min} > 0$.

The log-likelihood for a single observation is as follows:

$$\ell(\theta; d_i) = -\frac{1}{2}\ln\sigma^2 + \ln\Phi\left(\frac{-(y_i - x_i^\top\beta)\lambda_0}{\sigma}\right) - \frac{(y_i - x_i^\top\beta)^2}{2\sigma^2}, \tag{A1}$$

where $\sigma = \sqrt{\sigma^2}$.

The gradient of the log-likelihood is as follows:

$$\begin{aligned}
\frac{\partial\ell}{\partial\beta} &= \frac{\lambda_0}{\sigma}\cdot\frac{\phi(z)}{\Phi(z)}x_i + \frac{(y_i - x_i^\top\beta)}{\sigma^2}x_i,\\
\frac{\partial\ell}{\partial\lambda_0} &= \frac{\phi(z)}{\Phi(z)}\cdot\frac{-(y_i - x_i^\top\beta)}{\sigma},\\
\frac{\partial\ell}{\partial\sigma^2} &= -\frac{1}{2\sigma^2} - \frac{\phi(z)}{\Phi(z)}\cdot\frac{z}{2\sigma^2} + \frac{(y_i - x_i^\top\beta)^2}{2(\sigma^2)^2},
\end{aligned}$$

where $z = \frac{-(y_i - x_i^\top\beta)\lambda_0}{\sigma}$ and $\phi(z)$ is the standard normal probability density function.

The ℓ_1-norm of the gradient is given as follows:

$$\|\nabla\ell(\theta; d_i)\|_1 = \left\|\frac{\partial\ell}{\partial\beta}\right\|_1 + \left|\frac{\partial\ell}{\partial\lambda_0}\right| + \left|\frac{\partial\ell}{\partial\sigma^2}\right|.$$

Appendix A.1. Bounding the Components

Let $\varepsilon_i = y_i - x_i^\top \beta$ be the residual term. Given the data constraints $\|x_i\|_\infty \leq 1$ and $|y_i| \leq 1$, and the parameter constraint $\|\beta\|_1 \leq B$, the residual is bounded as follows:

$$\begin{aligned} |\varepsilon_i| &= |y_i - x_i^\top \beta| \\ &\leq |y_i| + |x_i^\top \beta| \\ &\leq 1 + \|x_i\|_\infty \|\beta\|_1 \\ &\leq 1 + B. \end{aligned}$$

Thus, the argument z is bounded by $|z| = \frac{|\varepsilon_i|\lambda_0}{\sigma} \leq \frac{(1+B)\lambda_{\max}}{\sigma_{\min}} := Z_{\max}$.

A crucial component is the inverse Mills ratio, $\frac{\phi(z)}{\Phi(z)}$. We establish its boundedness via the following lemma.

Lemma A1 (Mills Ratio Bound)**.** *For $z \in [-Z_{\max}, Z_{\max}]$, there exists a finite constant $M > 0$ such that the following is true:*

$$\frac{\phi(z)}{\Phi(z)} \leq M.$$

Proof. We consider the following two cases:

1. **Case 1:** $z \geq 0$. Since $\Phi(z) \geq \frac{1}{2}$ and $\phi(z) \leq \frac{1}{\sqrt{2\pi}}$, we have the following:

$$\frac{\phi(z)}{\Phi(z)} \leq \frac{1/\sqrt{2\pi}}{1/2} = \frac{2}{\sqrt{2\pi}}.$$

2. **Case 2:** $z < 0$. Using Mill's inequality, for $t > 0$, we have $1 - \Phi(t) \geq \frac{\phi(t)}{t+\sqrt{t^2+4}}$. For $z = -t$ with $t > 0$, we have $\Phi(z) = 1 - \Phi(t)$, so the following holds true:

$$\frac{\phi(z)}{\Phi(z)} = \frac{\phi(t)}{1-\Phi(t)} \leq \frac{\phi(t)}{\phi(t)/(t+\sqrt{t^2+4})} = t + \sqrt{t^2+4} \leq 2t + 2 = 2|z| + 2.$$

Since $|z| \leq Z_{\max}$, we have $\frac{\phi(z)}{\Phi(z)} \leq 2Z_{\max} + 2$. Therefore, a finite upper bound for the inverse Mills ratio is given by $M = \max\left\{\frac{2}{\sqrt{2\pi}}, 2Z_{\max} + 2\right\}$. □

Appendix A.2. Final Bound

Using the bounds derived above, we can now bound each gradient component. Given $\|x_i\|_\infty \leq 1$, we have $\|x_i\|_1 \leq p$.

- **Gradient with respect to β:**

$$\begin{aligned} \left\| \frac{\partial \ell}{\partial \beta} \right\|_1 &\leq \frac{\lambda_0}{\sigma} \cdot \frac{\phi(z)}{\Phi(z)} \|x_i\|_1 + \frac{|\varepsilon_i|}{\sigma^2} \|x_i\|_1 \\ &\leq \frac{\lambda_{\max}}{\sigma_{\min}} \cdot M \cdot p + \frac{1+B}{\sigma_{\min}^2} \cdot p \\ &= p\left(\frac{\lambda_{\max} M}{\sigma_{\min}} + \frac{1+B}{\sigma_{\min}^2} \right). \end{aligned}$$

- **Gradient with respect to λ_0:**

$$\begin{aligned}\left|\frac{\partial \ell}{\partial \lambda_0}\right| &\leq \frac{\phi(z)}{\Phi(z)} \cdot \frac{|\varepsilon_i|}{\sigma} \\ &\leq M \cdot \frac{1+B}{\sigma_{\min}}.\end{aligned}$$

- **Gradient with respect to σ^2:**

$$\begin{aligned}\left|\frac{\partial \ell}{\partial \sigma^2}\right| &\leq \frac{1}{2\sigma^2} + \frac{\phi(z)}{\Phi(z)} \cdot \frac{|z|}{2\sigma^2} + \frac{\varepsilon_i^2}{2(\sigma^2)^2} \\ &\leq \frac{1}{2\sigma_{\min}^2} + M \cdot \frac{Z_{\max}}{2\sigma_{\min}^2} + \frac{(1+B)^2}{2\sigma_{\min}^4} \\ &= \frac{1}{2\sigma_{\min}^2} + M \cdot \frac{(1+B)\lambda_{\max}}{2\sigma_{\min}^3} + \frac{(1+B)^2}{2\sigma_{\min}^4}.\end{aligned}$$

Combining all components, the final upper bound G on the gradient's ℓ_1-norm is as follows:

$$\begin{aligned}G = p\left(\frac{\lambda_{\max} M}{\sigma_{\min}} + \frac{1+B}{\sigma_{\min}^2}\right) + M \cdot \frac{1+B}{\sigma_{\min}} \\ + \frac{1}{2\sigma_{\min}^2} + M \cdot \frac{(1+B)\lambda_{\max}}{2\sigma_{\min}^3} + \frac{(1+B)^2}{2\sigma_{\min}^4},\end{aligned} \tag{A2}$$

where $M = \max\left\{\frac{2}{\sqrt{2\pi}}, 2Z_{\max} + 2\right\}$ and $Z_{\max} = \frac{(1+B)\lambda_{\max}}{\sigma_{\min}}$. Since all parameters are bounded within the constrained space, we have $\|\nabla \ell(\theta; d_i)\|_1 \leq G < \infty$, ensuring a well-defined sensitivity bound for differential privacy.

Appendix A.3. Parameter Impact Analysis on Gradient Sensitivity

The gradient sensitivity is particularly influenced by the parameters λ_0 and σ^2. A quantitative analysis reveals that λ_0 impacts sensitivity through two primary mechanisms—the inverse Mills ratio and a direct multiplicative effect. Numerical analysis demonstrates that as λ_0 increases from 0.5 to 2.0, the gradient bound G increases superlinearly. The dominant contribution shifts from the σ^2 component to the β component as λ_0 increases, with a critical threshold near $\lambda_0 \approx 1.5$ where numerical instability becomes prominent.

The parameter σ^2 exhibits a more dramatic influence, primarily due to an inverse power–law relationship and a cascade effect. The gradient bound contains terms that grow with the inverse fourth power of σ, and small values of σ^2 amplify both the Mills ratio and direct terms. The interaction between λ_0 and σ^2 creates a multiplicative amplification effect.

Appendix A.4. Numerical Stability Analysis

Stability Metrics

To quantify numerical stability, we define a score based on the parameters' proximity to regions of high sensitivity.

Definition A1 (Numerical Stability Score)**.** *The stability score is defined as follows:*

$$S(\lambda_0, \sigma_{\min}) = \frac{1}{1+Z_{\max}} \cdot \min\left\{1, \frac{\sigma_{\min}^2}{(1+B)^2}\right\}. \tag{A3}$$

References

1. Kumbhakar, S.C.; Lovell, C.K. *Stochastic Frontier Analysis*; Cambridge University Press: Cambridge, UK, 2003.
2. Aigner, D.; Lovell, C.K.; Schmidt, P. Formulation and estimation of stochastic frontier production function models. *J. Econom.* **1977**, *6*, 21–37. [CrossRef]
3. Meeusen, W.; van Den Broeck, J. Efficiency estimation from Cobb-Douglas production functions with composed error. *Int. Econ. Rev.* **1977**, *18*, 435–444. [CrossRef]
4. Goldberger, A.S. The interpretation and estimation of Cobb-Douglas functions. *Econom. J. Econom. Soc.* **1968**, *36*, 464–472. [CrossRef]
5. Kim, H.Y. The translog production function and variable returns to scale. *Rev. Econ. Stat.* **1992**, *74*, 546–552. [CrossRef]
6. Kumbhakar, S.C. Estimation of profit functions when profit is not maximum. *Am. J. Agric. Econ.* **2001**, *83*, 1–19. [CrossRef]
7. Kumbhakar, S.C. Efficiency measurement with multiple outputs and multiple inputs. *J. Product. Anal.* **1996**, *7*, 225–255. [CrossRef]
8. Fan, G.F.; Peng, L.L.; Hong, W.C.; Sun, F. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing* **2016**, *173*, 958–970. [CrossRef]
9. Tsionas, M.G. A coherent approach to Bayesian data envelopment analysis. *Eur. J. Oper. Res.* **2020**, *281*, 439–448. [CrossRef]
10. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [CrossRef]
11. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, 4–7 March 2006; Proceedings 3; Springer: Berlin/Heidelberg, Germany, 2006; pp. 265–284.
12. Chaudhuri, K.; Monteleoni, C.; Sarwate, A.D. Differentially private empirical risk minimization. *J. Mach. Learn. Res.* **2011**, *12*, 1069–1109. [PubMed]
13. Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H.B.; Mironov, I.; Talwar, K.; Zhang, L. Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 308–318.
14. Chaudhuri, K.; Sarwate, A.; Sinha, K. Near-optimal differentially private principal components. In Proceedings of the Neural Information Processing Systems Conference, Lake Tahoe, NV, USA, 3–6 December 2012.
15. Talwar, K.; Guha Thakurta, A.; Zhang, L. Nearly optimal private lasso. In Proceedings of the NIPS'15: Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015.
16. Bassily, R.; Smith, A. Local, private, efficient protocols for succinct histograms. In Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing, Portland, OR, USA, 14–17 June 2015; pp. 127–135.
17. Dwork, C.; Feldman, V. Privacy-preserving prediction. In Proceedings of the Conference on Learning Theory, PMLR, Stockholm, Sweden, 6–9 July 2018; pp. 1693–1702.
18. Zhang, Z.; Hu, R. Byzantine-robust federated learning with variance reduction and differential privacy. In Proceedings of the 2023 IEEE Conference on Communications and Network Security (CNS), Orlando, FL, USA, 2–5 October 2023; pp. 1–9.
19. Triastcyn, A.; Faltings, B. Bayesian differential privacy for machine learning. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual Event, 13–18 July 2020; pp. 9583–9592.
20. Smith, A. Efficient, differentially private point estimators. *arXiv* **2008**, arXiv:0809.4794. [CrossRef]
21. Bassily, R.; Feldman, V.; Talwar, K.; Guha Thakurta, A. Private stochastic convex optimization with optimal rates. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019.
22. Wang, Y.X. Revisiting differentially private linear regression: Optimal and adaptive prediction & estimation in unbounded domain. *arXiv* **2018**, arXiv:1803.02596. [CrossRef]
23. Jaggi, M. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In Proceedings of the International Conference on Machine Learning, PMLR, Atlanta, GA, USA, 16–21 June 2013; pp. 427–435.
24. Jain, P.; Thakurta, A. Differentially private learning with kernels. In Proceedings of the International Conference on Machine Learning, PMLR, Atlanta, GA, USA, 16–21 June 2013; pp. 118–126.
25. Talwar, K.; Thakurta, A.; Zhang, L. Private empirical risk minimization beyond the worst case: The effect of the constraint set geometry. *arXiv* **2014**, arXiv:1411.5417.
26. Raff, E.; Khanna, A.; Lu, F. Scaling up differentially private lasso regularized logistic regression via faster frank-wolfe iterations. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 36349–36363.
27. Khanna, A.; Lu, F.; Raff, E.; Testa, B. Differentially private logistic regression with sparse solutions. In Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security, Copenhagen, Denmark, 30 November 2023; pp. 1–9.
28. Acharya, K.; Boenisch, F.; Naidu, R.; Ziani, J. Personalized differential privacy for ridge regression. *arXiv* **2024**, arXiv:2401.17127. [CrossRef]
29. Wang, D.; Xu, J. Differentially private empirical risk minimization with smooth non-convex loss functions: A non-stationary view. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 1182–1189.

30. Cai, Z.; Li, S.; Xia, X.; Zhang, L. Private estimation and inference in high-dimensional regression with fdr control. *arXiv* **2023**, arXiv:2310.16260. [CrossRef]
31. Dwork, C.; Roth, A. The algorithmic foundations of differential privacy. *Found. Trends® Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
32. Dwork, C.; Rothblum, G.N.; Vadhan, S. Boosting and differential privacy. In Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, Las Vegas, NV, USA, 23–26 October 2010; pp. 51–60.
33. Frank, M.; Wolfe, P. An algorithm for quadratic programming. *Nav. Res. Logist. Q.* **1956**, *3*, 95–110. [CrossRef]
34. Bassily, R.; Smith, A.; Thakurta, A. Private empirical risk minimization: Efficient algorithms and tight error bounds. In Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science, Philadelphia, PA, USA, 18–21 October 2014; pp. 464–473.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.