Linear Regression as a Non-Cooperative Game

Stratis Ioannidis* and Patrick Loiseau†

*Technicolor Palo Alto, CA, USA. Email: stratis.ioannidis@technicolor.com.
†EURECOM, France. Email: patrick.loiseau@eurecom.fr.

Abstract. Linear regression amounts to estimating a linear model that maps features (e.g., age or gender) to corresponding data (e.g., the answer to a survey or the outcome of a medical exam). It is a ubiquitous tool in experimental sciences. We study a setting in which features are public but the data is private information. While the estimation of the linear model may be useful to participating individuals, (if, e.g., it leads to the discovery of a treatment to a disease), individuals may be reluctant to disclose their data due to privacy concerns. In this paper, we propose a generic game-theoretic model to express this trade-off. Users add noise to their data before releasing it. In particular, they choose the variance of this noise to minimize a cost comprising two components: (a) a privacy cost, representing the loss of privacy incurred by the release; and (b) an estimation cost, representing the inaccuracy in the linear model estimate. We study the Nash equilibria of this game, establishing the existence of a unique non-trivial equilibrium. We determine its efficiency for several classes of privacy and estimation costs, using the concept of the price of stability. Finally, we prove that, for a specific estimation cost, the generalized least-square estimator is optimal among all linear unbiased estimators in our non-cooperative setting: this result extends the famous Aitken/Gauss-Markov theorem in statistics, establishing that its conclusion persists even in the presence of strategic individuals.

Keywords: Linear regression, Gauss-Markov theorem, Aitken theorem, privacy, potential game, price of stability

1 Introduction

The statistical analysis of personal data is a cornerstone of several experimental sciences, such as medicine and sociology. Studies in these areas typically rely on experiments, drug trials, or surveys involving human subjects. Data collection has also become recently a commonplace—yet controversial—aspect of the Internet economy: companies such as Google, Amazon and Netflix maintain and mine large databases of behavioral information (such as, e.g., search queries or past purchases) to profile their users and personalize their services. In turn, this has raised privacy concerns from consumer advocacy groups, regulatory bodies, as well as the general public.

The desire for privacy incentivizes individuals to lie about their private information–or, in the extreme, altogether refrain from any disclosure. For example, an individual may be reluctant to participate in a medical study collecting

biometric information, concerned that it may be used in the future to increase her insurance premiums. Similarly, an online user may not wish to disclose her ratings to movies if this information is used to infer, e.g., her political affiliation. On the other hand, a successful data analysis may also provide a utility to the individuals from which the data is collected. This is evident in medical studies: an experiment may lead to the discovery of a treatment for a disease, from which an experiment subject may clearly benefit. In the cases of commercial data mining, users may benefit both from overall service improvements, as well as from personalization. If such benefits outweigh privacy considerations, users may consent to the collection and analysis of their data, e.g., by participating in a clinical trial, completing a survey, or using an online service.

In this paper, we approach the above issues through a non-cooperative game, focusing on a statistical analysis task called *linear regression*. We consider the following formal setting. A set of individuals $i \in \{1, \dots, n\}$ participate in an experiment, in which they are about to disclose to a data analyst a private variable $y_i \in \mathbb{R}$–e.g., the answer to a survey or the outcome of a medical test. Each individual i is associated with a feature vector $\boldsymbol{x}_i \in \mathbb{R}^d$, capturing public information such as, e.g., age, gender, etc. The analyst wishes to perform *linear regression* over the data, i.e., compute a vector $\boldsymbol{\beta} \in \mathbb{R}^d$ such that:

$$y_i \approx \boldsymbol{\beta}^T \boldsymbol{x}_i, \qquad \text{for all } i \in \{1, \dots, n\}.$$

However, individuals *do not* reveal their true private variables to the analyst in the clear: instead, before reporting these values, they *first add noise*. In our examples above, such noise addition aims to protect against, e.g., future use of the individual's biometric data by an insurance company, or inference of her political affiliation from her movie ratings. The higher the variance of the noise an individual adds, the better the privacy that she attains, as her true private value is obscured. On the other hand, high noise variance may also hurt the accuracy of the analyst's estimate of $\boldsymbol{\beta}$, the linear model computed in aggregate across multiple individuals. As such, the individuals need to strike a balance between the privacy cost they incur through disclosure and the utility they accrue from accurate model prediction.

Our contributions can be summarized as follows.

(*i*) We model interactions between individuals as a non-cooperative game, in which each individual selects the variance level of the noise to add to her private variable strategically. An individual's decision minimizes a cost function comprising two components: (a) a *privacy cost*, that is an increasing function of the added noise variance, and (b) an *estimation cost*, that decreases as the accuracy of the analyst's estimation of $\boldsymbol{\beta}$ increases. Formally, the estimation cost increases with the covariance matrix of the estimate of $\boldsymbol{\beta}$, when this estimate is computed through a least-squares minimization.

(*ii*) We characterize the Nash equilibria of the above game. In particular, we show that the above setting forms a potential game. Moreover, under appropriate assumptions on the privacy and estimation costs, there exists a unique pure Nash equilibrium at which individual costs are bounded.

(*iii*) Armed with this result, we determine the game's efficiency, providing bounds for the price of stability for several cases of privacy and estimation costs.

(*iv*) Finally, we turn our attention to the analyst's estimation algorithm. We show that, among the class of unbiased, linear estimators, generalized least squares is the estimator that yields the most accurate estimate, at equilibrium. In a formal sense, this extends the Aitken theorem in statistics, which states that generalized least squares estimation yields minimal variance among linear unbiased estimators. Our result implies that this optimality persists even if individuals strategically choose the variance of their data.

The remainder of this paper is organized as follows. We present related work in Section 2. Section 3 contains a review of linear regression and the definition of our non-cooperative game. We characterize Nash equilibria in Section 4 and discuss their efficiency in Section 5. Our Aitken-type theorem is in Section 6, and our conclusions in Section 7. Due to space constraints, long proofs are relegated to our technical report [13].

2 Related Work

Perturbing a dataset before submitting it as input to a data mining algorithm has a long history in privacy-preserving data-mining (see, e.g., [7,29]). Independent of an algorithm, early research focused on perturbing a dataset prior to its public release [8, 28]. Perturbations tailored to specific data mining tasks have also been studied in the context of reconstructing the original distribution of the underlying data [1], building decision trees [1], clustering [22], and association rule mining [3]. We are not aware of any study of such perturbation techniques in a non-cooperative setting, where individuals add noise strategically.

The above setting differs from the more recent framework of ϵ-differential privacy [9, 15], which has also been studied from the prespective of mechanism design [10, 16, 21]. In differential privacy, noise is added to the *output* of a computation, which is subsequently publicly released. Differential privacy offers a strong guarantee: changing an individual's input alters the distribution of the perturbed output at most by an $\exp\epsilon \approx 1 + \epsilon$ factor. The analyst performing the computation is a priori trusted; as such, individuals submit unadulterated inputs. In contrast, the classic privacy-preserving data-mining setting we study here assumes an untrusted analyst, which motivates input perturbation.

In experimental design [4, 24], an analyst observes the public features of a set of experiment subjects, and determines which experiments to conduct with the objective of learning a linear model. The quality of an estimated model is quantified through a scalarization of its variance [5]. Though many such scalarizations exist, we focus here on non-negative scalarizations, to ensure meaningful notions of efficiency (as determined by the price of stability, *c.f.* Section 5).

Several papers study problems of statistical inference from the perspective of mechanism design. Horel *et al.* [12] study a version of the experimental design problem in which subjects report their private values truthfully, but may lie about the costs they require for their participation. Closer to our setting, Dekel

et al. [6] consider a broad class of regression problems in which participants may misreport their private values, and determine loss functions under which empirical risk minimization is group strategy-proof–the special case of linear regression is also treated, albeit in a more restricted setting, in [23]. Our work differs in considering noise addition as a non-cooperative game, and studying the efficiency of its Nash equilibria, rather than mechanism design issues.

Our model has analogies to models used in *public good* provision problems (see, e.g., [20] and references therein). Indeed, the estimate variance reduction can be seen as a public good in that, when an individual contributes her data, all other individuals in the game benefit. Moreover, the perturbation technique used in our proof of Theorem 6 is similar to techniques used in public good models introduced in the context of traffic congestion [17, 18].

3 Model Description

In this section, we give a detailed description of our linear regression game and the players involved. Before discussing strategic considerations, we give a brief technical review of linear models, as well as key properties of least squares estimators; all related results presented here are classic (see, e.g., [11]).

Notational conventions. We use boldface type (e.g., $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{\beta}$) to denote vectors (all vectors are column vectors), and capital letters (e.g., A, B, V) to denote matrices. As usual, we denote by $S_+^d, S_{++}^d \subset \mathbb{R}^{d\times d}$ the sets of (symmetric) positive semidefinite (PSD) and positive definite matrices of size $d \times d$, respectively. For two positive semidefinite matrices $A, B \in S_+^d$, we write that $A \succeq B$ if $A-B \in S_+^d$; recall that $\succeq$ defines a partial order over S_+^d. We say that $F : S_+^d \to \mathbb{R}$ is non-decreasing in the positive semidefinite order if $F(A) \geq F(A')$ for any two $A, A' \in S_+^d$ such that $A \succeq A'$. Moreover, we say that a matrix-valued function $F : \mathbb{R}^n \to S_+^d$ is *matrix convex* if $\alpha F(\boldsymbol{\lambda}) + (1-\alpha)F(\boldsymbol{\lambda}') \succeq F(\alpha\boldsymbol{\lambda} + (1-\alpha)\boldsymbol{\lambda}')$ for all $\alpha \in [0,1]$ and $\boldsymbol{\lambda}, \boldsymbol{\lambda}' \in \mathbb{R}^n$. Given a square matrix $A = [a_{ij}]_{1\leq i,j\leq d} \in \mathbb{R}^{d\times d}$, we denote by $\mathrm{trace}(A)$ its *trace* (i.e., the sum of its diagonal elements), and by $\|A\|_F$ its *Frobenious norm* (i.e., the ℓ_2-norm of its d^2 elements).

3.1 Linear Models

Consider a set of n individuals, denoted by $N \equiv \{1, \cdots, n\}$. Each individual $i \in N$ is associated with a vector $\boldsymbol{x}_i \in \mathbb{R}^d$, the *feature vector*, which is public; for example, this vector may correspond to publicly available demographic information about the individual, such as age, gender, etc. Each $i \in N$ is also associated with a private variable $y_i \in \mathbb{R}$; for example, this may express the likelihood that this individual contracts a disease, the concentration of a substance in her blood or an answer that she gives to a survey.

Throughout our analysis, we assume that the individual's private variable y_i is a linear function of her public features $\boldsymbol{x}_i$. In particular, there exists a vector

$\boldsymbol{\beta} \in \mathbb{R}^d$, the *model*, such that the private variables are given by

$$y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i, \quad \text{for all } i \in N, \tag{1}$$

where the "inherent noise" variables $\{\epsilon_i\}_{i \in N}$ are i.i.d. zero-mean random variables in $\mathbb{R}$ with finite variance σ^2. We stress that we make no further assumptions on the noise; in particular, we *do not* assume it is Gaussian.

An analyst wishes to observe the y_i's and infer the model $\boldsymbol{\beta} \in \mathbb{R}^d$. This type of inference is ubiquitous in experimental sciences, and has a variety of applications. For example, the magnitude of $\boldsymbol{\beta}$'s coordinates captures the effect that features (e.g., age or weight) have on y_i (e.g., the propensity to get a disease), while the sign of a coordinate captures positive or negative correlation. Knowing $\boldsymbol{\beta}$ can also aid in prediction: an estimate of private variable $y \in \mathbb{R}$ of a new individual with features $\boldsymbol{x} \in \mathbb{R}^d$ is given by the inner product $\boldsymbol{\beta}^T \boldsymbol{x}$.

We note that the linear relationship between y_i and $\boldsymbol{x}_i$ expressed in (1) is in fact quite general. For example, the case where $y_i = f(\boldsymbol{x}) + \epsilon_i$, where f is a polynomial function of degree 2, reduces to a linear model by considering the transformed feature space whose features comprise the monomials $x_{ik}x_{ik'}$, for $1 \leq k, k' \leq d$. More generally, the same principle can be applied to reduce to (1) any function class spanned by a finite set of basis functions over $\mathbb{R}^d$ [11].

3.2 Generalized Least Squares Estimation

We consider a setup in which the individuals intentionally *perturb* or *distort* their private variable by adding excess noise. In particular, each $i \in N$ computes $\tilde{y}_i = y_i + z_i$ where z_i is a zero-mean random variable with variance σ_i^2; we assume that $\{z_i\}_{i \in N}$ are independent, and are also independent of the inherent noise variables $\{\epsilon_i\}_{i \in N}$. Subsequently, each individual reveals to the analyst (a) the perturbed variable $\tilde{y}_i$ and (b) the variance σ_i^2. As a result, the aggregate variance of the reported value is $\sigma^2 + \sigma_i^2$.

In turn, having access to the perturbed variables $\tilde{y}_i$, $i \in N$, and the corresponding variances, the analyst estimates $\boldsymbol{\beta}$ through *generalized least squares* (GLS) estimation. For $i \in N$, let $\lambda_i \equiv \frac{1}{\sigma^2 + \sigma_i^2}$ be the inverse of the aggregate variance. Denote by $\boldsymbol{\lambda} = [\lambda_i]_{i \in N}$ the vector of inverses and by $\Lambda = \mathrm{diag}(\boldsymbol{\lambda})$ the diagonal matrix whose diagonal is given by vector $\boldsymbol{\lambda}$. Then, the generalized least squares estimator is given by:

$$\hat{\boldsymbol{\beta}}_{\texttt{GLS}} = \operatorname*{arg\,min}_{\boldsymbol{\beta} \in R^d} \left(\sum_{i \in N} \lambda_i (\tilde{y}_i - \boldsymbol{\beta}^T \boldsymbol{x}_i)^2 \right) = (X^T \Lambda X)^{-1} X^T \Lambda \, \tilde{\boldsymbol{y}} \tag{2}$$

where $\tilde{\boldsymbol{y}} = [\tilde{y}_i]_{i \in N}$ is the n-dimensional vector of perturbed variables, and $X = [\boldsymbol{x}_i^T]_{i \in N} \in \mathbb{R}^{n \times d}$ the $n \times d$ matrix whose rows comprise the transposed feature vectors. Throughout our analysis, we assume that $n \geq d$ and that X has rank d.

Note that $\tilde{\boldsymbol{y}} \in \mathbb{R}^n$ is a random variable and as such, by (2), so is $\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$. It can be shown that $\mathbb{E}(\hat{\boldsymbol{\beta}}_{\texttt{GLS}}) = \boldsymbol{\beta}$ (i.e., $\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$ is unbiased), and

$$V(\boldsymbol{\lambda}) \equiv Cov(\hat{\boldsymbol{\beta}}_{\texttt{GLS}}) = \mathbb{E}\left[(\hat{\boldsymbol{\beta}}_{\texttt{GLS}} - \boldsymbol{\beta})^T (\hat{\boldsymbol{\beta}}_{\texttt{GLS}} - \boldsymbol{\beta}) \right] = (X^T \Lambda X)^{-1}.$$

The covariance V captures the uncertainty of the estimation of $\boldsymbol{\beta}$. The matrix

$$A(\boldsymbol{\lambda}) \equiv X^T \Lambda X = \sum_{i \in N} \lambda_i \boldsymbol{x}_i \boldsymbol{x}_i^T$$

is known as the *precision* matrix. It is positive semidefinite, i.e., $A(\boldsymbol{\lambda}) \in S_+^d$, but it may not be invertible: this is the case when $\operatorname{rank}(X^T \Lambda) < d$, i.e., the vectors $\boldsymbol{x}_i$, $i \in N$, for which $\lambda_i > 0$, do not span $\mathbb{R}^d$. Put differently, if the set of individuals providing useful information does not include d linearly independent vectors, there exists a direction $\boldsymbol{x} \in \mathbb{R}^d$ that is a "blind spot" to the analyst: the analyst has no way of predicting the value $\boldsymbol{\beta}^T \boldsymbol{x}$. In this degenerate case the number of solutions to the least squares estimation problem (2) is infinite, and the covariance is not well-defined (it is infinite in all such directions $\boldsymbol{x}$). Note however that, since X has rank d (and hence $X^T X$ is invertible), the set of $\boldsymbol{\lambda}$ for which the precision matrix is invertible is non-empty. In particular, it contains $(0, 1/\sigma^2]^n$ since $A(\boldsymbol{\lambda}) \in S_{++}^d$ if $\lambda_i > 0$ for all $i \in N$.

3.3 User Costs and a Non-Cooperative Game

The perturbations z_i are motivated by privacy concerns: an individual may be reluctant to grant unfettered access to her private variable or release it in the clear. On the other hand, it may be to the individual's advantage that the analyst learns the model $\boldsymbol{\beta}$. In our running medical example, learning that, e.g., a disease is correlated to an individual's weight or her cholesterol level may lead to a cure, which in turn may be beneficial to the individual.

We model the above considerations through cost functions. Recall that the action of each individual $i \in N$ amounts to choosing the noise level of the perturbation, captured by the variance $\sigma_i^2 \in [0, \infty]$. For notational convenience, we use the equivalent representation $\lambda_i = 1/(\sigma^2 + \sigma_i^2) \in [0, 1/\sigma^2]$ for the action of an individual. Note that $\lambda_i = 0$ (or, equivalently, infinite variance σ_i) corresponds to no participation: in terms of estimation through (2), it is as if this perturbed value is not reported.

Each individual $i \in N$ chooses her action $\lambda_i \in [0, 1/\sigma^2]$ to minimize her cost

$$J_i(\lambda_i, \lambda_{-i}) = c_i(\lambda_i) + f(\boldsymbol{\lambda}), \tag{3}$$

where we use the standard notation λ_{-i} to denote the collection of actions of all players but i. The cost function $J_i : \mathbb{R}_+^n \to \mathbb{R}_+$ of player $i \in N$ comprises two non-negative components. We refer to the first component $c_i : \mathbb{R}_+ \to \mathbb{R}_+$ as the *privacy cost*: it is the cost that the individual incurs on account of the privacy violation sustained by revealing the perturbed variable. The second component is the *estimation cost*, and we assume that it takes the form $f(\boldsymbol{\lambda}) = F(V(\boldsymbol{\lambda}))$, if $A(\boldsymbol{\lambda}) \in S_{++}^d$, and $f(\boldsymbol{\lambda}) = \infty$ otherwise. The mapping $F : S_{++}^d \to \mathbb{R}_+$ is known as a *scalarization* [5]. It maps the covariance matrix $V(\boldsymbol{\lambda})$ to a scalar value $F(V(\boldsymbol{\lambda}))$, and captures how well the analyst can estimate the model $\boldsymbol{\beta}$. The estimation cost $f : \mathbb{R}_+^n \to \bar{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{\infty\}$ is the so-called *extended-value extension* of $F(V(\boldsymbol{\lambda}))$: it equals $F(V(\boldsymbol{\lambda}))$ in its domain, and $+\infty$ outside its domain. Throughout our analysis, we make the following two assumptions:

Assumption 1 *The privacy costs $c_i : \mathbb{R}_+ \to \mathbb{R}_+$, $i \in N$, are twice continuously differentiable, non-negative, non-decreasing and strictly convex.*

Assumption 2 *The scalarization $F : S^d_{++} \to \mathbb{R}_+$ is twice continuously differentiable, non-negative, non-constant, non-decreasing in the positive semidefinite order, and convex.*

The monotonicity and convexity assumptions in Assumptions 1 and 2 are standard and natural. Increasing λ_i (i.e., decreasing the noise added by the individual) leads to a higher privacy cost. In contrast, increasing λ_i can only decrease the estimation cost: this is because decreasing the noise of an individual also decreases the variance in the positive semidefinite sense (as the matrix inverse is a PSD-decreasing function). Note that it is possible to relax Assumptions 1 and 2 (in particular, amend the twice-continuous differentiability assumption) without affecting most of our results, at the expense of an increased technical complexity in our proofs. We therefore focus on the above two assumptions for the sake of simplicity.

As a consequence of Assumption 2, the extended-value extension f is convex. The convexity of $F(V(\cdot))$ follows from the fact that it is the composition of the non-decreasing convex function $F(\cdot)$ with the matrix convex function $V(\cdot)$; the latter is convex because the matrix inverse is matrix convex. Moreover, f is twice continuously differentiable on its effective domain $\{\boldsymbol{\lambda} \in \mathbb{R}^n_+ : A(\boldsymbol{\lambda}) \in S^d_{++}\}$.

Scalarizations of positive semidefinite matrices and, in particular, of the covariance matrix $V(\boldsymbol{\lambda})$, are abundant in statistical inference literature in the context of experimental design [4, 5, 24]. We give two examples we use in our analysis below:

$$F_1(V) = \operatorname{trace}(V), \qquad F_2(V) = \|V\|_F^2. \tag{4}$$

Both scalarizations satisfy Assumption 2.

We denote by $\Gamma = \langle N, [0, 1/\sigma^2]^n, (J_i)_{i \in N} \rangle$ the game with set of players $N = \{1, \cdots, n\}$, where each each player $i \in N$ chooses her action λ_i in her action set $[0, 1/\sigma^2]$ to minimize her cost $J_i : [0, 1/\sigma^2]^n \to \mathbb{R}_+$, given by (3). We refer to a $\boldsymbol{\lambda} \in [0, 1/\sigma^2]^n$ as a *strategy profile* of the game Γ. We analyze the game as a *complete information game*, i.e., we assume that the set of players, the action sets and utilities are known by all players.

4 Nash Equilibria

We begin our analysis by characterizing the Nash equilibria of the game Γ. Observe first that Γ is a potential game [19]. Indeed, define the function $\Phi : [0, 1/\sigma^2]^n \to \bar{\mathbb{R}}$ such that

$$\Phi(\boldsymbol{\lambda}) = f(\boldsymbol{\lambda}) + \sum_{i \in N} c_i(\lambda_i), \quad (\boldsymbol{\lambda} \in [0, 1/\sigma^2]^n). \tag{5}$$

Then for every $i \in N$ and for every $\lambda_{-i} \in [0, 1/\sigma^2]^{n-1}$, we have

$$J_i(\lambda_i, \lambda_{-i}) - J_i(\lambda_i', \lambda_{-i}) = \Phi(\lambda_i, \lambda_{-i}) - \Phi(\lambda_i', \lambda_{-i}), \quad \forall \lambda_i, \lambda_i' \in [0, 1/\sigma^2]. \quad (6)$$

Therefore, Γ is a potential game with potential function Φ.

In the game Γ, each player chooses her contribution λ_i to minimize her cost. A Nash equilibrium (in pure strategy) is a strategy profile $\boldsymbol{\lambda}^*$ satisfying

$$\lambda_i^* \in \arg\min_{\lambda_i} J_i(\lambda_i, \lambda_{-i}^*), \qquad \text{for all } i \in N.$$

From (6), we see that (as for any potential game) the set of Nash equilibria coincides with the set of local minima of function Φ.

First note that there may exist Nash equilibria $\boldsymbol{\lambda}^*$ for which $f(\boldsymbol{\lambda}^*) = \infty$. For instance, if $d \geq 2$, $\boldsymbol{\lambda}^* = 0$ is a Nash equilibrium. Indeed, in that case, no individual has an incentive to deviate since a single $\lambda_i > 0$ still yields a non-invertible precision matrix $A(\boldsymbol{\lambda})$. In fact, any profile $\boldsymbol{\lambda}$ for which $A(\boldsymbol{\lambda})$ is non-invertible, and remains so under unilateral deviations, constitutes an equilibrium.

We call such Nash equilibria (at which the estimation cost is infinite) *trivial*. Existence of trivial equilibria can be avoided in practice using slight model adjustments. For instance, one can impose a finite upper bound on the variance σ_i of an individual i (or, equivalently, a positive lower bound on λ_i). Alternatively, the existence of d non-strategic individuals whose feature vectors span $\mathbb{R}^d$ is also sufficient to enforce a finite covariance at all $\boldsymbol{\lambda}$ across strategic individuals.

In the remainder, we focus on the more interesting *non-trivial* equilibria. Using the potential game structure of Γ, we derive the following result.

Theorem 1. *There exists a unique non-trivial equilibrium of the game Γ.*

Proof. Recall that the set of Nash equilibria coincides with the set of local minima of function Φ. To conclude the proof, we show that there exists a unique local minimum $\boldsymbol{\lambda}$ of Φ in the effective domain of f.

First note that, by Assumption 1, the privacy cost $c_i(\cdot)$ is finite on $[0, 1/\sigma^2]$ since it is continuous on a compact set. Therefore, $\Phi(\cdot)$ is finite *iif* $f(\cdot)$ is finite i.e., $\operatorname{dom}\Phi \equiv \{\boldsymbol{\lambda} : \Phi(\boldsymbol{\lambda}) < \infty\} = \operatorname{dom} f$, where dom is the effective domain. Recall that since X has rank d, $(0, 1/\sigma^2]^n \subset \operatorname{dom}\Phi$, and $\operatorname{dom}\Phi$ is non-empty.

By Assumptions 1 and 2, function Φ is strictly convex on its effective domain. Therefore it has at most one local minimum in $\operatorname{dom}\Phi$. Since $\operatorname{dom}\Phi$ is not compact, we still need to show that the minimum is achieved. By Assumption 1, the privacy cost derivatives are bounded and increasing. Let $M = \max_{i \in N} c'(1/\sigma^2) < \infty$ be the largest possible privacy cost derivative across all users and all λ_i's. On the other hand, the partial derivatives of the estimation cost can be written as

$$\begin{aligned}\frac{\partial f}{\partial \lambda_i}(\boldsymbol{\lambda}) &= -\operatorname{trace}\left(\frac{\partial F}{\partial V} \cdot (X^T \Lambda X)^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^T (X^T \Lambda X)^{-1}\right) \\ &= -\boldsymbol{x}_i^T (X^T \Lambda X)^{-1} \cdot \frac{\partial F}{\partial V} \cdot (X^T \Lambda X)^{-1} \boldsymbol{x}_i,\end{aligned}$$

where

$$\frac{\partial F}{\partial V} = \begin{pmatrix} \frac{\partial F}{\partial V_{11}} & \cdots \\ \vdots & \ddots \end{pmatrix}.$$

Hence, since F is non-constant, there exists an $i \in N$ for which $\frac{\partial f}{\partial \lambda_i}(\boldsymbol{\lambda})$ is unbounded. Therefore, it is possible to define $\bar{\mathcal{E}} \equiv \{\boldsymbol{\lambda} \in [0, 1/\sigma^2]^n : f(\boldsymbol{\lambda}) > K\}$ with K large enough so that $\max_{i \in N} \left|\frac{\partial f}{\partial \lambda_i}(\boldsymbol{\lambda})\right| > M$ for all $\boldsymbol{\lambda} \in \bar{\mathcal{E}}$. Let $\mathcal{E}$ be the complement of $\bar{\mathcal{E}}$ in $[0, 1/\sigma^2]^n$. Let $\boldsymbol{\lambda} \in \bar{\mathcal{E}}$. Since $\max_{i \in N} \left|\frac{\partial f}{\partial \lambda_i}(\boldsymbol{\lambda})\right| > M$, there exists $\boldsymbol{\lambda}' \in \mathcal{E}$ such that $\Phi(\boldsymbol{\lambda}') < \Phi(\boldsymbol{\lambda})$. Therefore, there exists a point in $\mathcal{E}$ for which Φ is smaller than anywhere outside $\mathcal{E}$. Finally, by Assumption 2, $\mathcal{E}$ is compact. We deduce that Φ has a unique minimum on $\operatorname{dom} \Phi$ which concludes the proof. □

The potential game structure of Γ has another interesting implication: if individuals start from an initial strategy profile $\boldsymbol{\lambda}$ such that $f(\boldsymbol{\lambda}) < \infty$, the so called *best-response dynamics* converge towards the unique non-trivial equilibrium (see, e.g., [26]). This implies that the non-trivial equilibrium is the only equilibrium reached when, e.g., all users start with non-infinite noise variance.

5 Price of Stability

Having established the uniqueness of a non-trivial equilibrium in our game, we turn our attention to issues of efficiency. We define the *social cost* function $C : \mathbb{R}^n \to \mathbb{R}_+$ as the sum of all individual costs, and say that a strategy profile $\boldsymbol{\lambda}^{\mathtt{opt}}$ is *socially optimal* if it minimizes the social cost, i.e.,

$$C(\boldsymbol{\lambda}) = \sum_{i \in N} c_i(\lambda_i) + n f(\boldsymbol{\lambda}), \qquad \text{and} \qquad \boldsymbol{\lambda}^{\mathtt{opt}} \in \operatorname*{arg\,min}_{\boldsymbol{\lambda} \in [0, 1/\sigma^2]^n} C(\boldsymbol{\lambda}).$$

Let $\mathtt{opt} = C(\boldsymbol{\lambda}^{\mathtt{opt}})$ be the minimal social cost. We define the *price of stability* (*price of anarchy*) as the ratio of the social cost of the best (worst) Nash equilibrium in Γ to $\mathtt{opt}$, i.e.,

$$\mathtt{PoS} = \min_{\boldsymbol{\lambda} \in \mathtt{NE}} \frac{C(\boldsymbol{\lambda})}{\mathtt{opt}}, \qquad \text{and} \qquad \mathtt{PoA} = \max_{\boldsymbol{\lambda} \in \mathtt{NE}} \frac{C(\boldsymbol{\lambda})}{\mathtt{opt}},$$

where $\mathtt{NE} \subset [0, 1/\sigma^2]^n$ is the set of Nash equilibria of Γ.

Clearly, in the presence of trivial equilibria, the price of anarchy is infinity. We thus turn our attention to determining the price of stability. Note however that since the non-trivial equilibrium is unique (Theorem 1), the price of stability and the price of anarchy coincide under slight model adjustments discussed in Section 4 that avoid existence of the trivial equilibria.

The fact that our game admits a potential function has the following immediate consequence (see, e.g., [26, 27]):

Theorem 2. *Under Assumptions 1 and 2,* $\mathtt{PoS} \leq n$.

Proof. Under Assumptions 1 and 2, the unique non-trivial equilibrium $\boldsymbol{\lambda}^*$ minimizes the potential function $\Phi(\boldsymbol{\lambda}) = \sum_{i\in N} c_i(\lambda_i) + f(\boldsymbol{\lambda})$. Then, for $\boldsymbol{\lambda}^{\texttt{opt}}$ a minimizer of the social cost:

$$\Phi(\boldsymbol{\lambda}^*) \leq \Phi(\boldsymbol{\lambda}^{\texttt{opt}}) = \sum_{i\in N} c_i(\lambda_i^{\texttt{opt}}) + f(\boldsymbol{\lambda}^{\texttt{opt}}) \leq \sum_{i\in N} c_i(\lambda_i^{\texttt{opt}}) + nf(\boldsymbol{\lambda}^{\texttt{opt}}) = \texttt{opt}$$

by the positivity of f. On the other hand, $C(\boldsymbol{\lambda}^*) \leq n\Phi(\boldsymbol{\lambda}^*)$, by the positivity of c_i, and the theorem follows. □

Improved bounds can be obtained for specific estimation and privacy cost functions. In what follows, we focus on the two inference cost functions given by (4). We make use of the following lemma, whose proof is in our technical report [13]:

Lemma 1. *If $A(\boldsymbol{\lambda}) \in S_{++}^d$, then for any $i \in N$,*

$$\frac{\partial \operatorname{trace}\left(A^{-1}(\boldsymbol{\lambda})\right)}{\partial \lambda_i} = -\boldsymbol{x}_i^T A^{-2}(\boldsymbol{\lambda})\boldsymbol{x}_i, \quad \textit{and} \quad \frac{\partial \|A^{-1}(\boldsymbol{\lambda})\|_F^2}{\partial \lambda_i} = -2\boldsymbol{x}_i^T A^{-3}(\boldsymbol{\lambda})\boldsymbol{x}_i.$$

We begin by providing a bound on the price of stability when privacy costs are monomial functions, proved in our technical report [13]. The following theorem characterizes the PoS in these cases, improving on the linear bound of Theorem 2:

Theorem 3. *Assume that the cost functions are given by $c_i(\lambda) = c_i\lambda^k$, where $c_i > 0$ and $k \geq 1$. If the estimation cost is given by the extended-value extension of $F_1(V) = \operatorname{trace}(V)$, then $\texttt{PoS} \leq n^{\frac{1}{k+1}}$. If the estimation cost is given by the extended-value extension of $F_2(V) = \|V\|_F^2$, then $\texttt{PoS} \leq n^{\frac{2}{k+2}}$.*

The proof of Theorem 3 relies on characterizing explicitly the socially optimal profile under relaxed constraints, and showing it equals the Nash equilibrium $\boldsymbol{\lambda}^*$ multiplied by a scalar. Moreover, the theorem states that, among monomial privacy costs, the largest PoS is $n^{\frac{1}{2}}$ for $F = F_1$, and $n^{\frac{2}{3}}$ for $F = F_2$. Both are attained at linear privacy costs; in fact, the above "worst-case" bounds can be generalized to a class of functions beyond monomials.

Theorem 4. *Assume that for every $i \in N$ the privacy cost functions $c_i : \mathbb{R}_+ \to \mathbb{R}_+$ satisfy Assumption 1. If the estimation cost is the extended-value extension of $F_1(V) = \operatorname{trace}(V)$, and the derivatives c_i' satisfy*

$$nc_i'(\lambda) \leq c_i'(n^{\frac{1}{2}}\lambda) \tag{7}$$

then $\texttt{PoS} \leq n^{\frac{1}{2}}$. Similarly, if the estimation cost is the extended-value extension of $F_2(V) = \|V\|_F^2$, and the derivatives c_i' satisfy

$$nc_i'(\lambda) \leq c_i'(n^{\frac{1}{3}}\lambda) \tag{8}$$

then $\texttt{PoS} \leq n^{\frac{2}{3}}$.

Theorem 4, proved in our technical report [13], applies to privacy cost functions that have the "strong" convexity properties (7) and (8). Roughly speaking, such functions grow no slower than cubic and fourth-power monomials, respectively. In contrast to Theorem 3 , in the case of Theorem 4, we cannot characterize the social optimum precisely; as a result, the proof relies on Brouwer's fixed point theorem to relate $\boldsymbol{\lambda}^{\texttt{opt}}$ to the non-trivial Nash equilibrium $\boldsymbol{\lambda}^*$.

We note that a similar worst-case efficiency of linear functions among convex cost families has also been observed in the context of other games, including routing [25] and resource allocation games [14]. As such, Theorems 3 and 4 indicate that this behavior emerges in our linear regression game as well.

6 An Aitken-Type Theorem for Nash Equilibria

Until this point, we have assumed that the analyst uses the generalized least-square estimator (2) to estimate model $\boldsymbol{\beta}$. In the non-strategic case, where $\boldsymbol{\lambda}$ (and, equivalently, the added noise variance) is fixed, the generalized least-square estimator is known to satisfy a strong optimality property: the so-called Aitken/Gauss-Markov theorem, which we briefly review below, states that it is the best linear unbiased estimator, a property commonly refered to as BLUE. In this section, we give an extension of this theorem, in the strategic case where $\boldsymbol{\lambda}^*$ is not a priori fixed, but is the equilibrium reached by users, itself depending on the estimator used by the analyst.

For all technical results in this section, we restrict ourselves to the case where $F(V) = F_1(V) = \text{trace}(V)$.

6.1 Linear Unbiased Estimators and the Aitken Theorem

A *linear* estimator $\hat{\boldsymbol{\beta}}_L$ of the model $\boldsymbol{\beta}$ is a linear map of the perturbed variables $\tilde{\boldsymbol{y}}$; i.e., it is an estimator that can be written as $\hat{\boldsymbol{\beta}}_L = L\tilde{\boldsymbol{y}}$ for some matrix $L \in \mathbb{R}^{d\times n}$. A linear estimator is called *unbiased* if $\mathbb{E}[L\tilde{\boldsymbol{y}}] = \boldsymbol{\beta}$ (the expectation taken over the inherent and added noise variables). Recall by (2) that the generalized least-square estimator $\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$ is an unbiased linear estimator with $L = (X^T\Lambda X)^{-1}X^T\Lambda$ and covariance $Cov(\hat{\boldsymbol{\beta}}_{\texttt{GLS}}) = (X^T\Lambda X)^{-1}$.

Any linear estimator $\hat{\boldsymbol{\beta}}_L = L\tilde{\boldsymbol{y}}$ can be written without loss of generality as

$$L = (X^T\Lambda X)^{-1}X^T\Lambda + D^T \tag{9}$$

where $D = (L - (X^T\Lambda X)^{-1}X^T\Lambda)^T \in \mathbb{R}^{n\times d}$. It is easy to verify that $\hat{\boldsymbol{\beta}}_L$ is unbiased if and only if $D^TX = 0$; in turn, using this result, the covariance of any linear unbiased estimator can be shown to be

$$Cov(\hat{\boldsymbol{\beta}}_L) = (X^T\Lambda X)^{-1} + D^T\Lambda^{-1}D \succeq Cov(\hat{\boldsymbol{\beta}}_{\texttt{GLS}}).$$

In other words, the covariance of the generalized least-square estimator is minimal in the positive-semidefinite order among the covariances of *all linear*

unbiased estimators. This optimality result is known as the Aitken theorem [2]. Applied specifically to homoschedastic noise (i.e., when all noise variances are identical), it is known as the Gauss-Markov theorem [11], which establishes the optimality of the ordinary least squares estimator. Both theorems provide a strong argument in favor of using least squares to estimate $\boldsymbol{\beta}$, in the presence of fixed noise variance.

6.2 Extension to a Non-Cooperative Game

Suppose now that the data analyst uses a linear unbiased estimator $\hat{\boldsymbol{\beta}}_L$ of the form (9), with a given matrix $D \in \mathbb{R}^{n\times d}$, which may depend on X. As before, we can define a game Γ in which each individual i chooses her λ_i to minimize her cost; this time, however, the estimation cost depends on the variance of $\hat{\boldsymbol{\beta}}_L$. A natural question to ask is the following: it is possible that, despite the fact that the analyst is using an estimator that is "inferior" to $\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$ in the BLUE sense, an equilibrium reached under $\hat{\boldsymbol{\beta}}_L$ is *better* than the equilibrium reached under $\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$? If so, despite the Aitken theorem, the data analyst would clearly have an incentive to use $\hat{\boldsymbol{\beta}}_L$ instead.

In this section, we answer this question in the negative, in effect extending Aitken's theorem to the case of strategic individuals. Formally, we consider the game $\Gamma = \langle N, [0, 1/\sigma^2]^n, (J_i)_{i\in N}\rangle$ defined as in Section 3.3, except that the estimation cost is the extended-value extension of $F_1(V(\boldsymbol{\lambda}))$ with

$$V(\boldsymbol{\lambda}) \equiv (X^T \Lambda X)^{-1} + D^T \Lambda^{-1} D, \qquad (\Lambda = \operatorname{diag} \boldsymbol{\lambda}). \tag{10}$$

Γ is still a potential game with potential function given by (5). Moreover, Assumption 2 still holds since $V(\cdot)$ given by (10) is a matrix convex function, and the extended-value extension $f(\cdot)$ is still convex.

Since the proof of Theorem 1 relied on the convexity of the potential, a straightforward adaptation of the proof gives the following result.

Theorem 5. *For any matrix $D \in \mathbb{R}^{n\times d}$, there exists a unique non-trivial equilibrium of the game Γ under the corresponding linear unbiased estimator* (9).

As for the case of `GLS`, this result follows from the uniqueness of a minimizer of the potential function attained in the effective domain.

We are now ready to state our extension of Aitken Theorem, proved in our technical report [13].

Theorem 6. *The generalized least-square estimator gives an optimal covariance among linear unbiased estimators, in the strategic case, in the order given by the scalarization F_1 used in the estimation cost. That is, for any linear unbiased estimator $\hat{\boldsymbol{\beta}}_L$, we have*

$$f(\boldsymbol{\lambda}^*_L) \geq f(\boldsymbol{\lambda}^*_{\texttt{GLS}}),$$

where $\boldsymbol{\lambda}^*_L$ and $\boldsymbol{\lambda}^*_{\texttt{GLS}}$ are the non-trivial equilibria for the linear unbiased estimator and for the generalized least-square estimator respectively.

Theorem 6 therefore establishes the optimality of $\hat{\boldsymbol{\beta}}_{\texttt{GLS}}$ amongst linear unbiased estimators w.r.t. the scalarization F_1, in the presense of strategic individuals. The proof uses perturbative techniques similar to the ones used in [18].

7 Concluding Remarks

This paper studies linear regression in the presence of cost-minimizing individuals, modeling noise addition as a non-cooperative game. We establish existence of a unique non-trivial Nash equilibrium, and study its efficiency for several different classes of privacy and estimation cost functions. We also show an extension of the Aitken/Gauss-Markov theorem to this non-cooperative setup.

The efficiency result in Theorem 3 gives specific bounds on the price of stability for monomial privacy costs. These bounds are sub-linear. However, the efficiency result in Theorem 4 indicates that a sub-linear price of stability can be attained for a much wider class of privacy cost functions. Nevertheless, Theorem 3 includes functions not covered by Theorem 4, which leaves open the question of extending the bounds of Theorem 4, potentially to all privacy costs satisfying Assumption 1. Moreover, both of these theorems, as well as Theorem 6, are shown for specific scalarizations of the estimator variance. Going beyond these scalarizations is also an interesting open problem.

Our Aitken/Gauss-Markov-type theorem is weaker than these two classical results in two ways. First, the optimality of the generalized least squares estimator is shown w.r.t. the partial order imposed by the scalarization F_1, rather than the positive semidefinite order. It would be interesting to strenghthen this result not only in this direction, but also in the case of the order imposed by other scalarizations used as estimation costs. Second, Theorem 6 applies to linear estimators whose difference from `GLS` does not depend on the actions $\boldsymbol{\lambda}$. In the presence of arbitrary dependence on $\boldsymbol{\lambda}$, the non-trivial equilibrium need not be unique (or even exist). Understanding when this occurs, and proving optimality results in this context, also remains open.

Finally, our model assumes that the variance added by each individual is known to the analyst. Amending this assumption brings issues of truthfulness into consideration: in particular, an important open question is whether there exists an estimator (viewed as a mechanism) that induces truthful noise reporting among individuals, at least in equilibrium. Again, an Aitken-type theorem seems instrumental in establishing such a result.

References

1. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: ACM SIGMOD International Conference on Management of Data. pp. 439–450 (2000)
2. Aitken, A.C.: On least squares and linear combinations of observations. Proceedings of the Royal Society of Edinburgh (1935)
3. Atallah, M., Bertino, E., Elmagarmid, A., Ibrahim, M., Verykios, V.: Disclosure limitation of sensitive rules. In: Workshop on Knowledge and Data Engineering Exchange (KDEX'99). pp. 45–52 (1999)
4. Atkinson, A., Donev, A., Tobias, R.: Optimum experimental designs, with SAS. Oxford University Press New York (2007)
5. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004)

6. Dekel, O., Fischer, F., Procaccia, A.D.: Incentive compatible regression learning. Journal of Computer and System Sciences 76(8), 759–777 (2010)
7. Domingo-Ferrer, J.: A survey of inference control methods for privacy-preserving data mining. In: Privacy-preserving data mining, pp. 53–80. Springer (2008)
8. Duncan, G.T., Mukherjee, S.: Optimal disclosure limitation strategy in statistical databases: Deterring tracker attacks through additive noise. Journal of the American Statistical Association 95(451), 720–729 (2000)
9. Dwork, C.: Differential privacy. In: International Colloquium on Automata, Languages and Programming (ICALP). pp. 1–12 (2006)
10. Ghosh, A., Roth, A.: Selling privacy at auction. In: ACM EC. pp. 199–208 (2011)
11. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, second edn. (2009)
12. Horel, T., Ioannidis, S., Muthukrishnan, S.: Budget feasible mechanisms for experimental design. arXiv preprint arXiv:1302.5724 (2013)
13. Ioannidis, S., Loiseau, P.: Linear regression as a non-cooperative game (2013), technical report, arXiv:1309.7824
14. Johari, R., Tsitsiklis, J.N.: Efficiency loss in a network resource allocation game. Mathematics of Operations Research 29(3), 407–435 (2004)
15. Kifer, D., Smith, A., Thakurta, A.: Private convex empirical risk minimization and high-dimensional regression. JMLR W&CP (Proceedings of COLT 2012) 23, 25.1–25.40 (2012)
16. Ligett, K., Roth, A.: Take it or Leave it: Running a Survey when Privacy Comes at a Cost. In: WINE. pp. 378–391 (2012)
17. Loiseau, P., Schwartz, G., Musacchio, J., Amin, S., Sastry, S.S.: Congestion pricing using a raffle-based scheme. In: NetGCoop (October 2011)
18. Loiseau, P., Schwartz, G., Musacchio, J., Amin, S., Sastry, S.S.: Incentive mechanisms for internet congestion management: Fixed-budget rebate versus time-of-day pricing. IEEE/ACM Transactions on Networking (2013), to appear
19. Monderer, D., Shapley, L.S.: Potential games. Games and Economic Behavior 14(1), 124–143 (1996)
20. Morgan, J.: Financing public goods by means of lotteries. Review of Economic Studies 67(4), 761–84 (October 2000)
21. Nissim, K., Smorodinsky, R., Tennenholtz, M.: Approximately optimal mechanism design via differential privacy. In: Innovations in Theoretical Computer Science (ITCS). pp. 203–213 (2012)
22. Oliveira, S.R., Zaiane, O.R.: Privacy preserving clustering by data transformation. In: SBBD. pp. 304–318 (2003)
23. Perote, J., Perote-Pena, J.: Strategy-proof estimators for simple regression. Mathematical Social Sciences 47(2), 153–176 (2004)
24. Pukelsheim, F.: Optimal design of experiments, vol. 50. Society for Industrial Mathematics (2006)
25. Roughgarden, T., Tardos, E.: How bad is selfish routing? Journal of the ACM 49(2), 236–259 (Mar 2002)
26. Sandholm, W.H.: Population Games and Evolutionary Dynamics. MIT Press (2010)
27. Schäfer, G.: Online social networks and network economics (2011), lecture notes, Sapienza University of Rome
28. Traub, J.F., Yemini, Y., Woźniakowski, H.: The statistical security of a statistical database. ACM Transactions on Database Systems (TODS) 9(4), 672–679 (1984)
29. Vaidya, J., Clifton, C.W., Zhu, Y.M.: Privacy Preserving Data Mining. Springer (2006)