arXiv:2601.10626v1 [math.ST] 15 Jan 2026

纵向线性回归的差分隐私推断

GETOAR SOPA、MARCO AVELLA MEDINA 和 CYNTHIA RUSH

Abstract. 差分隐私（DP）为发布统计量提供了严谨的框架，同时保护数据集中包含的个人信息。尽管差分隐私线性回归已取得重大进展，现有方法几乎都只针对项目级差分隐私设定，即每位用户仅贡献一条观测。许多科学和经济应用则涉及纵向数据或面板数据，其中每位用户都会贡献多条相互依赖的观测。在这些设定下，项目级差分隐私无法提供充分保护，而用户级差分隐私——保护个人的完整轨迹——才是恰当的隐私概念。我们为用户级差分隐私下的纵向线性回归估计与推断构建了一个全面框架。我们提出了一种基于局部回归聚合的用户级私有回归估计量，并在短程依赖条件下建立了有限样本保证和渐近正态性。对于推断，我们提出了一种隐私化偏差校正协方差估计量，它能够自动满足异方差和自相关一致性。这些结果为存在依赖时纵向线性回归中的实用用户级差分隐私估计与推断提供了首个统一框架，并具有坚实的理论保证和良好的实证表现。

1. Introduction

差分隐私（DP）是一种通用框架，可为数据被用于计算发布统计量的用户提供最坏情况下的隐私保证（Dwork，2006年）。尽管差分隐私最初主要是一种理论构想，但如今已广泛应用于公共、商业和研究领域（Abowd，2018年；Cormode等人，2018年；Liu等，2024年）。差分隐私的核心思想是，必须向非平凡统计量中注入经过适当校准的噪声，才能确保发布结果具有隐私性。因此，构建差分隐私统计量的首要挑战，是设计随机化机制，在提供有意义的隐私保证的同时，尽可能降低统计效率的损失。

关键词和短语。 差分隐私、纵向线性回归、异方差和自相关一致推断。哥伦比亚大学统计系。本研究部分由美国国家科学基金会资助，项目编号为 DMS-2310973（Avella Medina）和 DMS-2413828（Rush）。

线性回归或许是应用统计学中最重要的工具。因此，针对线性回归的差分隐私估计与推断受到广泛研究，也就不足为奇了（Sheffet，2019年；Barrientos 等，2019年；Cai 等，2021年；Alabi 等，2022年；Milionis 等，2022年；Alabi 和 Vadhan，2023年；Amin 等，2023年；Cai 等，2023年；Barrientos 等，2024年）。迄今为止，在差分隐私约束下对线性回归的研究，都是在项目级差分隐私设定下开展的：每位用户恰好贡献一个数据点，而差分隐私随机化旨在隐藏任何一位用户所贡献的数据点。然而，在许多应用中，每位用户都会为线性回归问题贡献多个数据点，例如医学实验（Flynn 等，2009年；Church 等，2010年）、社会科学研究（Sobel，2012年；Mewes 等，2021年）以及微观经济数据分析（Jacobson 等，1993年；Autor 等，2014年）。我们将这种回归称为纵向线性回归或面板线性回归。由于此类数据集中通常存在时间依赖性，独立同分布数据模型并不适用于这些应用。

此外，在纵向线性回归等统计应用中，项目级差分隐私无法提供令人满意的隐私保证。直观来说，除非同时屏蔽某位用户的所有数据点，否则我们不会认为该用户对某个统计量的贡献是私密的。近年来，许多差分隐私领域的研究通过引入一种称为用户级（或个人级）差分隐私（uDP）的框架来应对这一问题，该框架可防范单个用户全部数据贡献发生最坏情况变化所造成的影响（Levy 等，2021年；Cummings 等，2022年；Ghazi 等，2023年；Kent 等，2024年；Agarwal 等，2025年）。尽管用户级差分隐私也是适用于纵向数据的自然隐私框架，但据我们所知，现有关于用户级差分隐私的研究都假设数据独立同分布，而如上所述，这通常并不是纵向回归的良好模型。近期有一项研究同时考虑了用户级差分隐私和依赖关系，是个例外；该研究由本文一位作者完成（Roth 和 Avella-Medina，2025年），分析了在由对数 Sobolev 不等式刻画的依赖结构下用户级差分隐私均值估计的收敛速率。然而，该论文没有讨论推断，因此其中的程序主要具有理论意义，并非出于实际应用的考虑。

在本研究中，我们研究了用户级差分隐私约束和时间依赖性下线性回归模型中的估计与推断。本文的主要贡献总结如下：

(a) **自适应用户级差分隐私均值估计：**我们提出了一种新颖的算法，只需对数据生成过程有最低限度的了解，即可进行均值估计。这在纵向回归场景中尤为重要，原因有二。首先，该算法实用性强，可直接应用于真实的回归问题，且无需

事先掌握过多信息。其次，纵向数据往往具有显著的时间依赖性，因此传统的裁剪加噪算法调参策略并不适用。我们的实证结果表明，对于中等样本量，该算法具有出色的性能。我们将Karwa 和 Vadhan(2018年)；Biswas 等(2020年的方法拓展到一般的、可能存在相关性的 数据。

(b) 用户级差分隐私线性回归估计：我们提出了一种用户级隐私估计量，用于估计纵向线性回归系数。该估计量基于对各用户数据分别计算的线性回归估计值取平均。我们认为，在用户级差分隐私环境下，这种估计量通常优于基于汇总统计量隐私化的估计量；并证明了在用户数据存在短期依赖的情况下，该隐私估计量具有有限样本和渐近收敛性。

(c) **用户级差分隐私推断：**我们提出了用于估计隐私保护 OLS 估计量渐近协方差的用户级隐私估计量，其依据是经过隐私化处理的样本协方差估计量，并加入有限样本修正项。我们表明，该估计量具有一致性，且能自动实现异方差和自相关一致性，并展示了其在小样本中的准确性。我们的构造与聚类稳健标准误（Liang 和 Zeger，1986年；Arellano，1987年）有相似之处，但针对我们的隐私保护分布式环境进行了调整。

(d) 用户级差分隐私组间差异估计量： 我们提出了用于估计两组（例如处理组和对照组）之间系数差异及其渐近协方差的估计量，并证明了相应的收敛保证。这项贡献在实践中特别有意义，因为在应用中经常会遇到此类变量，例如处理状态、性别或族裔等分类变量。

(e) **依赖情形下的集中性结果：** 作为分析长期数据普通最小二乘法的副产品，我们在强混合条件下获得了多种新的集中不等式。具体而言，我们构造方法中的一个关键中间步骤，将 Hsu 等人 (2012) 对有限样本的分析推广到了自相关情形。为此，我们借鉴了 Merlev`ede 等人(2011) 的研究成果。我们的结果和技术也与近期关于短期依赖下回归的研究有些相似（Fan 等人，2016年）。我们认为，这些集中性结果本身就很有意义，并且很可能在未来的研究中找到更多有用的应用。

本文其余部分安排如下。第 2 节介绍用于阐述主要结果所需的差分隐私和概率论背景。第 3 节给出支撑纵向估计量的核心均值估计算法。该算法返回一个差分隐私均值估计量，后续各节将对其进行适当

实例化，以提供具有强统计保证的用户级差分隐私估计量。第 4 节介绍我们的主要隐私纵向线性回归估计量。第 5 节则补充介绍其渐近协方差的隐私估计量。我们利用这些估计量构造隐私近似枢轴量，以进行一致推断。第 6 节通过使用模拟数据的数值示例展示我们方法在样本量适中的情况下的实用性；第 7 节则使用儿童哮喘管理计划（CAMP，2000）教学数据集加以展示。附录包含所有证明。

2. Preliminaries

对于自然数 $K \in \mathbb{N}$，我们用 $[K]$ 表示集合 $\{1,...,K\}$。对于向量 $v \in \mathbb{R}^p$，我们记作 $\|v\| = \sqrt{v^\top v}$。对于实矩阵 $A \in \mathbb{R}^{p\times q}$，我们记 $\|A\| = \max_{v\in\mathbb{R}^d:\|v\|=1} \|Av\|$，并分别用 $\lambda_{\min}(A)$和 $\lambda_{\max}(A)$表示 A的最小和最大特征值。我们用 $\mathbb{M}_{symm}^{d\times d}$表示实对称 $d \times d$ 矩阵的空间。如果存在$\exists C > 0$ 使得对于 f 和 g的任意可容许参数，$f(\cdot) \le C(g(\cdot))$，则我们记 $f(x) = O(g(x))$。类似地，如果存在 $\exists c > 0$使得对于 f和 g的任意可容许参数，$f(\cdot) \ge cg(\cdot)$，则我们记$f(x) = \Omega(g(x))$；如果 $f(x) = O(g(x))$ 且 $f(x) = \Omega(g(x))$，则记$f(x) \asymp g(x)$。此外，如果$f(x)\, g(x) \to 0$ as $x \to \infty$，则我们记 $f(x) = o(g(x))$。最后，对于事件 E，我们用 $\mathbf{1}\{E\}$ 或 $\mathbb{1}_E$ 表示事件 E的示性函数。

2.1. 高斯差分隐私。

本文所维护的隐私度量是高斯差分隐私（GDP），它提供了一个易于自然解释的框架以及若干理想性质（Dong 等人，2022年）。为定义 GDP，我们必须首先介绍权衡函数的概念。

定义 2.1。 设 P 和 Q 是定义在同一可测空间（$\mathcal{X}$，$\mathcal{F}$）上的两个概率分布。一个（可能是随机化的）拒绝规则是任意 $\mathcal{F}$-可测函数$\phi : \mathcal{X} \mapsto [0,1]$。权衡函数 $T(P, Q) : [0,1] \mapsto [0,1]$ 定义为 $T(P, Q)(\alpha) = \inf\{1 - \mathbb{E}_Q[\phi] : \mathbb{E}_P[\phi] \le \alpha\}$，其中下确界取遍拒绝规则集合 ϕ。

直观而言，权衡函数衡量的是通过统计假设检验区分分布 P和 Q 的难度：对于任意给定的 $\alpha \in [0,1]$，如果在 P 下错误拒绝的概率不超过 α，它就量化了在 Q下错误拒绝概率的最小可能值。高斯差分隐私通过比较区分由两个相邻数据集生成的输出空间的难度与区分两个高斯分布的难度，来评估算法的隐私。为此，对于固定的隐私参数 $\mu > 0$，定义 $G_\mu := T(N(0, 1)\ N(\mu, 1))$，它满足 $G_\mu(\alpha) = \Phi(\Phi^{-1}(1-\alpha)-\mu)$，其中 $\alpha \in [0,1]$，而 $\Phi(\cdot)$ 表示标准正态累积分布函数。为了刻画用户级

高斯差分隐私（uGDP），我们首先介绍用户级差分隐私框架中的相邻数据集概念。在这里，数据集是一个矩阵 $X^n \in \mathbb{R}^{nT\times d}$，其中包含 T 个测量值；每个用户对应的测量值维度为 d，共有 n个用户。

定义 2.2。 设 $X^n = (X_1^\intercal, \ldots, X_n^\intercal)^\intercal$ 和 $X_i \in \mathbb{R}^{T\times d}$ 以及 $X'^n = (X_1'^\intercal, \ldots X_n'^\intercal)$ 构成一个大小相同的数据集。若 $d_H(X^n, X'^n) := \sum_{i=1}^n \mathbb{1}\{X_i \neq X_i'\} = 1$，则 X^n 和 X'^n 是用户级相邻的，其中 $\mathbb{1}\{X_i \neq X_i'\} = 0$ 当且仅当矩阵 X_i 和 X_i' 的所有元素都相等。

我们强调，若两个数据集仅在 n 个用户的数据上存在差异，则这两个数据集是用户级邻居；允许该用户的所有 T 个条目在两个数据集中都不同。

定义 2.3。 随机统计量 h 满足 μ-uGDP，如果对于任意两个用户级相邻数据集 X^n 和 X'^n，以及所有 $\alpha \in [0,1]$.，都有 $T(h(X^n)\ h(X'^n))(\alpha) \geq G_\mu(\alpha)$

因此，若一个随机统计量满足 μ-uGDP，则对于任意两个相邻输入，区分由它们生成的输出至少与基于单次抽样区分一个 $N(0,1)$和 $N(\mu,1)$ 一样困难。因此，这种差分隐私概念可以在假设检验框架下轻松理解（Wasserman 和 Zhou，2010年）。我们强调，权衡函数中所考虑的随机性来自统计量 h 的随机性质，而非数据 X^n；在此目的下，数据被视为固定。请注意，随着 μ 缩小，统计量 h 会变得更加隐私，因为区分变得更加困难。确保算法满足高斯差分隐私的典型方法是添加经过适当校准的高斯噪声。为形式化这一点，我们首先引入算法全局敏感度的概念。

定义 2.4。设 $h : \mathbf{X}^n \to \mathbb{R}^d$ 是一个统计量。那么， h 的用户级全局敏感度定义为 $\mathrm{GS}(h) = \sup_{d_H(X^n,\ X'^n)=1} \|h(X^n) - h(X'^n)\|$.

这使我们能够引入高斯机制及其隐私保证。

命题 2.5 (定理 1，出自Dong 等人（2022）)。 设 h 是用户级全局敏感度为 $GS(h)$ 的统计量。那么，随机统计量 $\tilde{h}(X^n) = h(X^n) + {}^{GS_{(h)}}{}_{\mu}\ Z$，其$Z \sim N(0, I_d)$，满足$\mu$-*uGDP*。

最后，我们利用 GDP 的以下两个重要性质，以推导更复杂算法的隐私保证。

命题 2.6 （Dong 等人（2022）的命题 4）。 设随机统计量 $h : \mathcal{X}^n \mapsto \mathcal{Y}_1$ 满足 μ-*uGDP*。此外，设 $g : \mathcal{Y}_1 \mapsto \mathcal{Y}_2$ 是一个从该统计量的输出空间 h 映射到空间 $\mathcal{Y}_2$的（可能是随机的）映射。那么 $g \circ h : \mathcal{X}^n \mapsto \mathcal{Y}_2$ 也满足 μ-*uGDP*。

直观地说，前述命题在 DP 文献中被称为后处理性质，其含义是，对隐私统计量的输出进行处理不会削弱其隐私保证。

命题 2.7（Dong 等人（2022）的推论 *2*）。对于 $j \in [N]$，设 h_j 为满足 μ_j-*uGDP* 的随机统计量。则它们的 N重组合满足 $\sqrt{\sum_{j=1}^{N} \mu_j^2}$-*uGDP*。

上述结果称为 GDP 的组合性质，使我们能够精确量化发布多个隐私化统计量所造成的隐私损失；这些统计量可能将先前隐私化的统计量作为输入。例如，如果随机统计量 $h : \mathrm{X}^n \mapsto \mathrm{Y}_1$ 满足 $\mu/\sqrt{2}$-GDP，且随机统计量 $g : \mathrm{X}^{n-1} \times \mathrm{Y}_1 \mapsto \mathrm{Y}_2$ 满足$\mu/\sqrt{2}$-GDP，那么联合输出 $h(X^n)$ 和 $g(X^n, h(X^n))$ 满足 $\sqrt{\mu^2/2 + \mu^2/2} = \mu$-GDP。命题 2.6 和 2.7 使我们能够使用简单的构建模块构造复杂的隐私算法，同时仍能提供有意义的隐私保证。

2.2. **概率论**。我们回顾本文中反复用到的一些基本概率概念，首先介绍次高斯随机向量。

定义 2.8。若随机向量 $X \in \mathbb{R}^d$ 满足对于所有 $u \in \mathbb{S}^{d-1}$ 和 $\lambda \in \mathbb{R}$，都有 $\mathbb{E}[e^{\lambda u^{\mathsf{T}}(X - \mathbb{E}[X])}] \leq \exp(\sigma_X^2 \lambda^2/2)$，则称其为 σ_X^2-次高斯。

接下来，我们定义本文中使用的随机过程类型。以下定义 2.9–2.12 是标准定义，改编自 Brockwell 和 Davis (1991年)。

定义 2.9。实值随机过程 $(\varepsilon_t)_{t \in \mathbb{Z}}$ 是定义在联合概率空间 $(\mathrm{X}, \mathrm{F}, \mathbb{P})$ 上的一组随机变量。若过程 $(\varepsilon_t)_{t \in \mathbb{Z}}$ 满足$\mathbb{E}[\varepsilon_t^2] < \infty$ 且对所有 $t \in \mathbb{Z}$都有 $\mathbb{E}[\varepsilon_t] = m$，并且进一步有 $\gamma(t, s) := \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = \gamma(t + r, s + r)$ 对所有 $t, s, r \in \mathbb{Z}$成立，则称其为弱平稳。

我们进一步定义强混合序列的概念。设 $(X_t)_{t \in \mathbb{Z}}$ 是$\mathbb{R}^d$中的一个过程，$\mathrm{F}_a^b = \sigma(X_a、\ldots\ldots、X_b)$ 是由 X_a 通过 X_b 诱导的 σ-代数。

定义 2.10。随机过程 $(X_t)_{t \in \mathbb{Z}}$ 的 α-混合系数定义为$\alpha(k) = \sup_{t \in \mathbb{Z}} \sup_{A \in \mathcal{F}_{-\infty}^t,\ B \in \mathcal{F}_{t+k}^{\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$，若 $\alpha(k) \to 0$ 当 $k \to \infty$时趋于零，则称该过程为强混合过程。此外，若存在$\exists C,\ c > 0$ 使得 $\forall k \geq 1$ 满足 $\alpha(k) \leq Ce^{-ck}$，则称该过程为指数强混合过程。

最后，我们所有的动机性示例都属于 ARMA过程，尤其是因果ARMA过程。直观地说，因果性保证误差或初始值的影响会随着时间推移而衰减。这些定义可以直接推广到多元时间序列。

定义 2.11. 弱平稳随机过程 $(\varepsilon_t)_{t\in\mathbb{Z}}$ 是 ARMA(p,q) 过程，当且仅当它可以表示为 $\varepsilon_t - \sum_{k=1}^{p}\phi_k\varepsilon_{t-k} = r_t + \sum_{l=1}^{q}\theta_l r_{t-l}$，其中 r_t 均值为零，且对于所有 t， $s \in \mathbb{Z}$，$\mathrm{Cov}(r_t, r_s) = \mathbb{1}_{t=s}\sigma^2$。定义多项式 $\phi(z) := 1 - \sum_{k=1}^{p}\phi_k z^k$ 和 $\theta(z) := 1 + \sum_{l=1}^{q}\theta_l z^l$ 后，ARMA过程可以表示为 $\phi(B)\varepsilon_t = \theta(B)r_t$，其中 B 是后移算子。我们假设 $\phi(\cdot)$ 和 $\theta(\cdot)$ 没有共同的零点。

定义 2.12。 如 2.11 中所定义的 ARMA(p,q) 过程 $(\varepsilon_t)_{t\in\mathbb{Z}}$ 称为因果过程，当且仅当 $\phi(z)$（定义 2.11 中所定义）的所有零点都严格位于单位圆之外。在这种情况下，存在绝对可和的系数 $(\psi_j)_{j\geq 0}$，使得 $\varepsilon_t = \sum_{j\geq 0}\psi_j r_{t-j}$ 对所有 $t \in \mathbb{Z}$.

3. Adaptive differentially private mean estimation

本文贯穿始终的主要工具是算法 1，这是一种自适应的高斯差分隐私均值估计算法。它在概念上与现有文献中的方法相似，例如（Biswas 等，2020年）提出的 Coinpress 算法及其先驱工作（Karwa 和 Vadhan，2018年； Kamath等人， 2019），但有一个重要区别：算法 1 不需要对底层数据生成过程作出显著的结构性假设，也不需要了解该过程。下一节将看到，它可以用于构造具有强统计保证的 uGDP估计量。

我们的均值估计算法如下。我们从一个半径较大的球开始，不失一般性地令其中心为 0。随后，算法通过迭代进行私有检验，判断是否有足够多的用户数据落在一个新球内。该球以先前的私有均值估计为中心，半径比前一个球小一个常数因子。如果是，算法便利用细化球内的数据估计新的私有均值；与前一个球相比，该球的敏感度有所降低。算法重复这一过程，直到私有检验表明有过多数据点落在新提出的球外；此时，算法会基于通过私有检验的最终球内数据点计算最终的私有估计量。从这个意义上说，我们的算法也可以被视为一种自动化、迭代式的 Propose-Test-Release 实现（Dwork 和 Lei，2009年；Avella-Medina 和 Brunel，2020年）。在本算法的具体实现中，我们对私有检验阈值进行配置，使得只要所有用户都落在某次迭代的球内，算法就以高概率不会在该次迭代终止。需要指出的是，如果我们只关心估计输入数据的均值，算法 1 只需输出 $\hat{m}_{\mathrm{DP}}$，即最终的均值估计。然而，算法 1 还会输出以下量： r^*、 B^* 和 $\hat{m}_{r^*-1}$，分别表示算法的终止轮次、最终噪声尺度和最终球心。这些额外的输出值将在第 5 节用于私有估计协方差矩阵。

算法 1 基于差分隐私修剪的均值估计 $\mathrm{DPTrimMean}((D_i)_{i=1}^n,\ \mu,\ B,\ R,\ \xi)$

输入： 数据 $(D_i)_{i=1}^n$，隐私参数 μ，半径 B，最大轮数 R，失败概率 ξ
设定： $\tau \leftarrow n - \frac{2}{\mu}\sqrt{2R\log\frac{4R}{\xi}}$ 与 $n_{LB} \leftarrow \max(2\tau - n,\ 1)$
初始化： $\hat{m}_{-2} \leftarrow 0, \hat{m}_{-1} \leftarrow 0, r \leftarrow 0, r^* \leftarrow -1$
当 $r \le R$ 且 $r^* = -1$ **执行**
  $\omega_i \leftarrow \mathbb{1}\{\|D_i - \hat{m}_{r-1}\| \le B/2^r\}$, $\omega \leftarrow \sum_{i=1}^n \omega_i + \frac{2\sqrt{R}}{\mu} Z_r$, $Z_r \sim \mathsf{N}(0,\ 1)$
  如果 $\omega < \tau$ **则**
    $S \leftarrow \{i : \|D_i - \hat{m}_{r-2}\| < B/2^{r-1}\}$, $n_{\tilde{S}} \leftarrow$ 最大值$(|S|, n_{LB})$, $\mathsf{C} \leftarrow \sqrt{2 - \frac{r}{R}}$
    $\hat{m}_{\mathsf{DP}} \leftarrow \hat{m}_{r-2} + \frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-2}) + \frac{2\sqrt{2}B}{2^{r-1}\mathsf{C}\mu n_{LB}} N_*$, $N_* \sim \mathsf{N}(\mathbf{0}, I)$
    $r^* \leftarrow r - 1$, $B^* \leftarrow \frac{2\sqrt{2}B}{2^{r^*}\mathsf{C}\mu n_{LB}}$
  否则，如果 $r = R$，**则**
    $S \leftarrow \{i : \|D_i - \hat{m}_{R-1}\| < B/2^R\}$, $n_{\tilde{S}} \leftarrow$ 最大值$(|S|, n_{LB})$
    $\hat{m}_{\mathsf{DP}} \leftarrow \hat{m}_{R-1} + \frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{R-1}) + \frac{2\sqrt{2}B}{2^R\mu n_{LB}} N_*$, $N_* \sim \mathsf{N}(\mathbf{0}, I)$
    $r^* \leftarrow R$, $B^* \leftarrow \frac{2\sqrt{2}B}{2^{r^*}\mu n_{LB}}$
  else
    $S \leftarrow \{i : \|D_i - \hat{m}_{r-1}\| < B/2^r\}$, $n_{\tilde{S}} \leftarrow$ 最大值$(|S|, n_{LB})$
    $\hat{m}_r \leftarrow \hat{m}_{r-1} + \frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-1}) + \frac{4B\sqrt{R}}{2^r\mu n_{LB}} N_r$, $N_r \sim \mathsf{N}(\mathbf{0}, I)$
    $r \leftarrow r + 1$
  结束条件判断
结束循环
输出空间： $(\hat{m}_{\mathsf{DP}}, \hat{m}_{r^*-1}, r^*, B^*)$

与现有方法相比，我们算法的主要实际优势在于，它能以数据驱动的方式适应用户数据中的集中程度。因此，截断和添加噪声的尺度都不会过大或过小。算法通过私有检验实现这一点，而不必像Karwa 和 Vadhan （2018年）；Kamath等人（2019）；Biswas 等（2020年）那样，依赖高斯尾部等结构性假设，或要求预先掌握协方差在常数因子范围内的估计。在这一构造中，除了预先确定隐私预算 μ外，还须提供三个调优参数，分别对应数据界限 B 、最大细化轮次 R 以及我们希望估计量在有限样本下达到的失败概率 ξ 。由于投影球的半径是在指数递减的网格上逐步探索的，实践中可以为初始半径选择较大且保守的 B 值。[1]

我们用一个简单的例子来说明我们的算法，其中 $D_i \overset{iid}{\sim} N([50, 50]^\intercal,\ 10I_2)$，对于 $i = 1, \ldots, 300$。我们使用隐私参数 $\mu = 1$、半径 $B = 10^{14}$、最大轮数 $R = 50$ 和失败概率 $\xi = 0.05$ 初始化算法。注意， B 设为

[1]事实上，如果在第一次迭代中 B 过小，可以增大参数 B, R，更新 $\mu \leftarrow \mu(1 - 1/(2\sqrt{R}))$，然后重新启动算法。为简单起见，我们不考虑这种情况。

远高于所需值，且足以容纳近乎无界的数据。算法在 $r^* = 43$ 次迭代后终止。虽然初始噪声估计值为 $\hat{}m_0 = [2.3 \cdot 10^{13}, 8.09 \cdot 10^{12}]^\intercal$，但最终输出空间 $\hat{}m_{\mathsf{DP}} = [49.93, 50.22]^\intercal$ 接近真实值 $[50, 50]^\intercal$。此外，最终使用的半径为 $\frac{B}{2^{r^*}} = 10^{14}/2^{43} = 11.37$，与理论上的一致集中半径非常接近。为此，请注意 $\|D_i - [50, 50]^\intercal\|^2/10 \sim \chi^2_2$，因此

$$\mathbb{P}\Big(\max_{i=1,\dots,300} \|D_i - [50, 50]^\intercal\| \le \sqrt{20 \log \frac{1}{1 - (1-\xi)^{1/300}}} \Big) = 1 - \xi,$$

对于 $\xi = 0.05$，该半径约为 13.17。我们的半径略小，因为我们允许少数用户落在投影球之外。然而，我们的算法能够在不事先了解数据分布的情况下，以差分隐私方式求得该半径。图 1 展示了算法的最后五个步骤。

关于算法 1 的主要结果在定理 3.1 中陈述。该结果涉及确定性数据，随机性完全来自算法注入的随机性。证明见附录B.1。

定理 3.1. 算法 *1* 满足 μ-高斯差分隐私。此外，假设存在 $B, K_1, K_2 > 0$ 以及常数 m，使得 最大值$_{i\in[n]} \|D_i\| \le B$, $\|D - m\| \le K_1$ 且 最大值$_{i\in[n]} \|D_i - m\| \le K_2$，其中 $\bar{D} := \frac{1}{n}\sum_{i=1}^n D_i$。若 $R \ge C$对数$\frac{B}{K_2}$且 $n \ge C'(\frac{1}{\mu}\sqrt{R(}$对数$\frac{R}{\xi} + d))$ 对于通用常数$C, C' > 0$成立，则算法 *1* 以至少 $1 - \xi$ 的概率满足，

$$\|\hat{m}_{\mathsf{DP}} - m\| = O\Big(K_1 + \frac{K_2}{n\mu}\sqrt{R \log \frac{R}{\xi}} + \frac{K_2}{n\mu}\sqrt{d + \log \frac{1}{\xi}}\Big) = O\Big(K_1 + \frac{K_2}{n\mu}\sqrt{d + R \log \frac{R}{\xi}}\Big).$$

通过利用集中性结果，并在用户数据充分集中的高概率事件上分析算法，可以轻松修改上述结果以适用于随机数据。下面，我们用一个典型的独立同分布次高斯用户级差分隐私均值估计问题来说明这一点。首先，我们指出这一设置下的以下标准集中性结果。

引理 3.2. 设 $w_{i,t} \overset{iid}{\sim}$次高斯$(\sigma^2)$ 满足 $i \in [n], t \in [T]$ 和 $m := \mathbb{E}[w_{i,t}] \in \mathbb{R}^d$。定义 $(\bar{w}_i)_{i=1}^n$，其中 $\bar{w}_i := \frac{1}{T}\sum_{t=1}^T w_{i,t}$。那么，对于绝对常数 $C > 0$，以至少 $1 - \xi$ 的概率，

$$\max_{i\in[n]} \|\bar{w}_i - m\| \le C\sigma\sqrt{\frac{1}{T}\Big(d + \log \frac{n}{\xi}\Big)} \quad \text{and} \quad \Big\|\frac{1}{n}\sum_{i=1}^n \bar{w}_i - m\Big\| \le C\sigma\sqrt{\frac{1}{nT}\Big(d + \log \frac{1}{\xi}\Big)}.$$

现在，我们可以通过分析引理 3.2 的 $1 - \xi$ 概率事件下的算法 1，将定理 3.1 应用于用户级隐私估计 m 的问题。具体而言，这使我们能够在定理 3.1 中取 $K_1 = O(\sqrt{\frac{1}{nT}(d + \log \frac{1}{\xi})})$ 和 $K_2 = O(\sqrt{\frac{1}{T}(d + \log \frac{n}{\xi})})$。由此可立即得到以下推论。

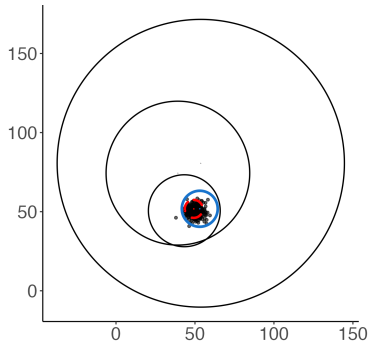


FIGURE 1. 算法 1 最后五次迭代的示意图。红色虚线圆圈表示最后一次尝试但被拒绝的均值细化，而蓝色圆圈表示最终投影球。

推论 3.3. 令 $w_{i,t} \overset{iid}{\sim}$次高斯$(\sigma^2)$满足 $i \in [n], t \in [T]$和 $m := \mathbb{E}[w_{i,t}] \in \mathbb{R}^d$。使用隐私参数 μ 将算法 *1* 应用于$(\bar{w}_i)_{i=1}^n$，其中 $\bar{w}_i := \frac{1}{T}\sum_{t=1}^T w_{i,t}$，可得到一个 μ-*uGDP* 估计值$\hat{m}_{\mathsf{DP}}$。此外，在满足定理 *3.1* 的缩放条件时，以至少 $1-\xi$ 的概率，

$$\|\hat{m}_{\mathsf{DP}} - m\| = O\Big(\sigma\sqrt{\frac{1}{nT}\Big(d + \log\frac{1}{\xi}\Big)} + \sigma\sqrt{\frac{1}{n^2\mu^2 T}\Big(R\log\frac{R}{\xi} + d\Big)\Big(d + \log\frac{n}{\xi}\Big)}\Big).$$

此处的 R对数 $\frac{R}{\xi}$ 项是我们为搜索正确阶数半径而进行的私有自适应搜索所付出的代价。具体而言，已知精确集中半径的“预言机”程序（例如 Levy 等 (2021年)中的算法 2）可得到类似的误差界，其中 R对数 $\frac{R}{\xi}$ 被 log $\frac{1}{\xi}$ 取代。在第 4 节中，我们将这些工具应用于用户级隐私线性回归系数估计问题。

3.1. **讨论。** 算法 1 通过迭代运行一系列敏感度为 1 的查询，并在其中一个查询低于阈值时终止，从而以隐私方式确定最优投影半径。这与 AboveThreshold 框架（例如 Dwork和Roth（2014）中的算法 1）有相似之处；该框架还会在 ϵ-差分隐私框架中对阈值添加噪声。采用这种做法，只需添加方差为常数阶的噪声，就足以实现固定的隐私保证，而不受可能执行的查询数量影响。在 μ-高斯差分隐私设置下，或许也可以进行类似的处理，从而有可能改进定理 3.1 中的误差界；但对此进行全面探讨超出了本文的范围。

4. DIFFERENTIALLY PRIVATE LONGITUDINAL LINEAR REGRESSION ESTIMATION

4.1. **模型设定与假设。** 我们在 uGDP 框架下考察的第一个、最基本的纵向线性回归模型为

$$y_{i,t} = \beta^{\mathsf{T}} x_{i,t} + \varepsilon_{i,t}, \tag{1}$$

其中，$i \in [n]$ 表示用户索引，$t \in [T]$ 表示时间索引，$x_{i,t}$ 是一组时变协变量，可能包含一个常数项以纳入截距。[2] 我们关注的系数是 β。下文中，我们常用 $Y \in \mathbb{R}^{nT}$ 表示由$y_{i,t}$堆叠而成的向量，用 $X \in \mathbb{R}^{nT \times d}$ 表示由$x_{i,t}$堆叠而成的矩阵，并用$\varepsilon \in \mathbb{R}^{nT}$ 表示由$\varepsilon_{i,t}$堆叠而成的向量。我们用 Y_i、X_i和 ε_i 表示用户 i所贡献的相应数据块。我们对协变量和误差分布提出以下正则条件。

假设 4.1.（指数混合设计）协变量满足$\Gamma_i^{-2} x_{i,t} \sim$ 次高斯(σ_x^2)，其中 $\mathbb{E}[x_{i,t} x_{i,t}^\top] = \Gamma_i \in \mathbb{R}^{d \times d}$ 具有 $d = O(1)$，且 $0 < c_\Gamma \le \lambda_{\min}(\Gamma_i) \le \lambda_{\max}(\Gamma_i) \le C_\Gamma < \infty$。此外，我们假设$(x_{i,t})_{t \in \mathbb{Z}}$ 是强混合 $\forall i \in [n]$ 的，且 $\alpha_{x_i}(k) \le C\exp\{-ck\}$，常数为 C，$c > 0$。最后，我们假设对于 $i \neq j$，x_i 与 x_j 相互独立。

示例 *1.* 设对所有 $i \in [n]$、$t \in [T]$，$x_{i,t} = (1; z_{i,t}^\top)^\top$。若对每个 $i \in [n]$，$(z_{i,t})_{t \in \mathbb{Z}}$服从因果多元*ARMA(p,q)*过程$z_{i,t} - m = \sum_{j=0}^{\infty} \Psi_j e_{i,t-j}$，其中 $\lambda_{\min}(\sum_{j=0}^{\infty} \Psi_j \Psi_j^\top) > 0$，且对$e_{i,t} \overset{iid}{\sim}$次高斯$(\sigma_x^2)$有如下条件：它们关于勒贝格测度具有密度 $\mathbb{E}[e_{i,t}] = 0$，$Var(e_{i,t}) =: \Sigma_e$，以及 $0 < \lambda_{\min}(\Sigma_e) \le \lambda_{\max}(\Sigma_e) < \infty$，则假设 *4.1* 成立。更多细节参见附录 *F.1*。

假设 4.2。（强混合误差）对于所有 $i \in [n]$ 和 $T \in \mathbb{N}$，误差满足$\mathbb{E}[\varepsilon_i] = 0$且 $\mathrm{Var}(\varepsilon_i) =: \Sigma_T \in \mathbb{R}^{T \times T}$，其中 $\lambda_{\min}(\Sigma_T) \ge c_\Sigma > 0$，并且对于 $i \neq j$，(ε_i, X_i) 与 (ε_j, X_j) 相互独立。此外，对于 $i \in [n]$，$\varepsilon_i \mid X_i \sim$次高斯$(\sigma_\varepsilon^2)$，其中 σ_ε^2不依赖于T。最后，$\mathbb{E}[\varepsilon \mid X] = 0$，$(\varepsilon_{i,t})_{t \in \mathbb{Z}}$和 $(x_{i,t}\varepsilon_{i,t})_{t \in \mathbb{Z}}$均为弱平稳序列，而 $(x_{i,t}, \varepsilon_{i,t})_{t \in \mathbb{Z}}$为强混合序列，其混合系数为 $\alpha_{\epsilon_i, x_i}(k) \le Ck^{-3-q}$，其中某些 $q > 0$ 和 $C > 0$。

在渐近分析中，我们考虑序列（n，T）$\to \infty$。因此，ε_i隐式地依赖于 T，但为简洁起见，我们省略这种依赖；在本文其余部分，我们通常也对 Σ_T作同样处理，简记为 Σ。需要指出的是，由于假设对所有 $T \ge 1$，$\varepsilon_i \mid X_i \sim \mathrm{subG}(\sigma_\varepsilon^2)$，因此 $\|\Sigma\| \le \sigma_\varepsilon^2$在 T上一致成立。

例2。如果对于每个 $i \in [n]$，ε_i（ε_i）$_{i=1}^n$ 由因果 *AR(p)* 过程 $\varepsilon_{i,t} = \sum_{j=0}^{\infty} \psi_j r_{i,t-j}$生成，其中 $r_{i,t} \overset{iid}{\sim}$次高斯（$\sigma_r^2$）的分布相对于勒贝格测度绝对连续，则满足假设 *4.2*。特别地，我们有 $\sigma_\varepsilon^2 \le \sigma_r^2 \sum_{j \ge 0} \psi_j^2$。更多细节请参见附录 *F.2*。

[2]如果每位用户都有各自唯一的截距，我们的结果不会改变，但保证仅适用于非截距系数。

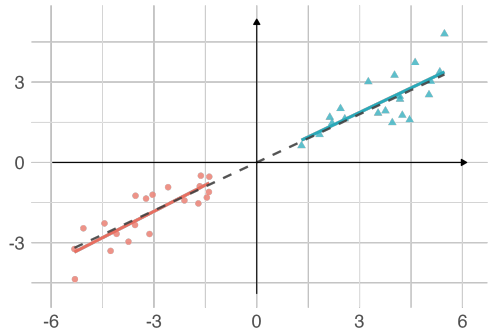


FIGURE 2. 两个用户的 Gram 矩阵并不一致集中。然而，它们的局部回归拟合值会紧密集中于共同的回归系数附近。

4.2. 用户级差分隐私回归估计。 假设 4.1 和 4.2 确保可以构造用户特定的 $\sqrt{T}$-一致估计量 β，其形式为

$$\hat{\beta}_i := X_i^{\dagger} Y_i, \tag{2}$$

其中，† 表示摩尔－彭罗斯伪逆。对于任意实矩阵 $A \in \mathbb{R}^{m \times p}$，它具有 $A^{\dagger} = (A^{\mathsf{T}} A)^{\dagger} A^{\mathsf{T}}$ 的性质；对于可逆矩阵 $B \in \mathbb{R}^{m \times m}$，则具有 $B^{\dagger} = B^{-1}$ 的性质。因此，当 $\mathrm{Rank}(X_i) = d$ 时，我们有 $\hat{\beta}_i = (X_i^{\mathsf{T}} X_i)^{-1} X_i^{\mathsf{T}} Y_i$，即通常的局部最小二乘估计量。此外，无论 X_i 的秩如何，式 (2) 中的估计量都具有如下性质：它是最小 ℓ_2-范数最小二乘解中的唯一解。由于 DPTRIMMEAN 接收数值输入，我们采用摩尔－彭罗斯局部最小二乘解，以便在 $\mathrm{Rank}(X_i) < d$ 时不泄露秩亏信息。利用式 (2) 中的局部估计量，中心最终可以通过用户级样本均值构造一个分布式的 $\sqrt{nT}$-一致但非私有估计量：

$$\bar{\beta} := \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_i. \tag{3}$$

因此，$\bar{\beta}$的一种自然的用户级隐私对应形式，也是我们在本文中采用的方法，是改为以差分隐私方式计算用户回归估计值的用户级样本均值。我们不妨思考一下，是什么使它成为一个理想的候选估计量，以及在许多情形下，为什么它可能优于聚合最小二乘估计量的差分隐私对应形式。在项目级差分隐私线性回归中，通常会分别计算差分隐私汇总统计量 $(X^{\mathsf{T}}X)_{\mathrm{DP}}$ 和 $(X^{\mathsf{T}}Y)_{\mathrm{DP}}$，然后将二者组合起来，得到 $\widehat{\beta_{\mathrm{SSDP}}} := \widehat{((X^{\mathsf{T}}X)_{\mathrm{DP}})^{-1}(X^{\mathsf{T}}Y)_{\mathrm{DP}}}$，例如参见 Wang（2018年）；Sheffet（2019）；Alabi 等（2022年）。然而，对于回归系数的汇总统计量，不太可能实现围绕共同均值的均匀集中性；而显然，从定理 3.1 可知，这种集中性对于用户级差分隐私估计是理想的。因此，平均估计量可能是更有吸引力的隐私化候选方案。图 2 是一个简单示例。此外，由于线性回归问题具有线性性质，而且不同于大多数 M 估计问题，

ˆ 当 X_i 满秩时，每个局部估计 β_i 都是无偏估计量，因此，平均局部估计是一种有效的替代方案，因为平均后不会残留 $O(\frac{1}{T})$ 偏差项。

现在，我们转而采用第 3 节中介绍的技术，对候选估计 (3) 进行隐私化。具体而言，关于这些估计值围绕真实回归系数 β的集中性，我们有如下结果；该结果与独立同分布情形下的结论相同，后者见（Hsu 等人，2012，定理11），但对 ˆT的要求略有提高。与 Hsu 等人（2012）一样，证明通过将差值 $\beta_i - \beta$ 写成次高斯 ε_i 的一个“有界”线性变换来进行，并且要求协变量的经验二阶矩矩阵性质良好的高概率事件成立。与他们结果的主要区别在于，在我们的研究中，协变量存在时间依赖性，因此必须同时控制多个格拉姆矩阵。该引理的证明见附录C.1。

ˆ　　¯**引理 4.3. </style}满足假设 4.1-4.2，并令如(2)**所示，　β 如(3)所示。则以至少　$1-\xi$ 的概率，

$$\max_{i\in[n]} \|\hat{\beta}_i - \beta\| = O\Big(\sigma_\varepsilon \max_{i\in[n]} \|\Gamma_i^{-\frac{1}{2}}\| \sqrt{\frac{1}{T}\big(d + \log\frac{n}{\xi}\big)}\Big),$$
$$\|\bar{\beta} - \beta\| = O\Big(\sigma_\varepsilon \sqrt{\frac{1}{nT}\big(\frac{1}{n}\sum_{i=1}^{n} \|\Gamma_i^{-\frac{1}{2}}\|^2\big)\big(d + \log\frac{1}{\xi}\big)}\Big),$$

当　$T \geq C(d + \log \frac{nT}{\xi})^2,$，其中　$C > 0$　是一个依赖于　$O(1)$ 总体参数的常数。

因此，前文的讨论以及第 3 节中的讨论表明，可以采用以下私有ˆ估计量来估计　β。首先，每位用户　$i \in [n]$ 估计其局部回归系数　β_i。然后，我们将 $(\beta_i)_{i=1}^n$ 作为输入提供给算法 1，并将所得输出记为　β_{DP}。根据算法 1 的构造，该估计值因此满足　μ-uGDP。在陈述估计量　β_{DP}的效用保证之前，我们先引入一些量作为结果成立的阈值；下文中，　$B_{\min}$ 和 $R_{\min}$ 与算法 1 的输入B 和 R 相关。对于某些常数　C_1，　C_2，　C_3，　$C_4 > 0$，它们可能依赖于　O(1) 总体参数，并且　$g(n,\ T,\ \xi,\ d)$:= $C_2(d + \log \frac{nT}{\xi})^2$:

—

$$n_{\min}(R,\mu,\xi,d) := C_1 \frac{1}{\mu}\sqrt{R\Big(\log\frac{R}{\xi} + d\Big)}, \quad T_{\min}(n,T,\xi,d) := \min\{T : T \geq g(n,T,\xi,d)\},$$
(4)
$$R_{\min}(T,B) := C_3 \log B\sqrt{T}, \quad \text{and} \quad B_{\min}(\beta,n,T,\xi,d) := \|\beta\| + C_4\sqrt{\frac{1}{T}\big(d + \log\frac{n}{\xi}\big)}.$$

下文中，我们通常省略这些函数对其自变量的依赖关系。

这些要素足以为我们最终得到的 β 隐私估计量提供高概率误差界。然而，要推导其渐近分布，我们要求用户的格拉姆矩阵具有一致良好的性质，具体形式如下。

假设 4.4. 对于给定的 $k \geq 1$，存在一个 $\delta > 0$ 和 $T_0 > d$，使得对于所有 $i \geq 1$，$X_i^\top X_i$ 几乎必然可逆，且 $\sup_{i\geq 1,\ T\geq T_0} \mathbb{E}[\|(\frac{1}{T}X_i^\top X_i)^{-1}\|^{k+\delta}] < \infty$.–

下面的引理给出了一个容易验证的充分条件，在该条件下假设 4.4 成立。证明见附录 C.2。

引理 4.5. 若对于任意 $v \in \mathbb{S}^{d-1}$，记 $w_{i,\ t} := v^\top x_{i,\ t}$，则假设 *4.4* 成立；其中 $W_i = (w_{i,1}, \ldots w_{i,\ T})$ 的密度 f_{W_i} 满足 上确界$_{i\geq 1,\ w\in\mathbb{R}^T} f_{W_i}(w) \leq K^T$，且存在一个与 n、T和 v无关的常数$K > 0$，并且对任意 $\delta' > 0$，取 $T_0 \geq 2(d + k + \delta')$。

直观而言，这一充分条件仅要求协变量的密度随 T 增长时不会增长过快；例如，在依赖性极强的情况下，该条件就成立。下面我们给出一个满足这一条件的简单示例。

示例 3. 若$(x_{i,\ t})_{t\geq 1}$ 由因果多元高斯 *AR(q)* 生成，且创新项的协方差矩阵为 $\Sigma_z \succ \mathbf{0}$，则该过程满足引理 *4.5* 的条件，因此也满足假设 *4.4*。更多详情请参见附录 *F.3*。

最后，我们的结果将用到以下用于估计量渐近正态性的增长条件：

$$\frac{1}{n\mu^2}\Big(R\log\frac{R}{\xi}\Big)\Big(\log\frac{n}{\xi}\Big) = o(1), \quad \text{and} \quad \xi = o(1). \tag{5}$$

现在我们可以陈述用户级差分隐私回归估计量的隐私和统计保证，见定理 4.6。需要指出的是，定理 4.6(b) 中的高概率误差界可直接由算法 1 的确定性保证（见定理 3.1）与引理 4.3 的集中性结论结合得到。证明见附录 C.3。

定理 4.6。设 β_i如(2) 所示，且令 $\beta_{\mathsf{DP}} = \mathrm{DPTrimMean}((\beta_i)_{i=1}^n, \mu, B, R, \xi/2)$。

(a) 隐私。 $\hat{\beta}_{\mathsf{DP}}$ 满足 μ-*uGDP*。

(b) 误差界。 在假设 *4.1-4.2* 下，若 $B \geq B_{\min}$, $R \geq R_{\min}$, $n \geq n_{\min}$ 和 $T \geq T_{\min}$ 使用 (4) 中的值，则以至少 $1 - \xi$ 的概率，

$$\|\hat{\beta}_{\mathsf{DP}} - \beta\| = O\Big(\sigma_\varepsilon\Big[\sqrt{\frac{\frac{1}{n}\sum_{i=1}^n \|\Gamma_i^{-\frac{1}{2}}\|^2(d + \log\frac{1}{\xi})}{nT}} + \max_{i\in[n]} \|\Gamma_i^{-\frac{1}{2}}\| \sqrt{\frac{(R\log\frac{R}{\xi} + d)(d + \log\frac{n}{\xi})}{n^2\mu^2 T}}\Big]\Big).$$

(c) 渐近正态性。 给定 $\Lambda_i := \sum_{\ell=-\infty}^{\infty} \mathbb{E}[x_{i,0}x_{i,\ell}^{\mathsf{T}}\varepsilon_{i,0}\varepsilon_{i,\ell}]$，假设存在 $V := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}$。若在 (5) 下，***(b)*** 中的假设仍然成立，并且满足假设 ***4.4***，且 $k = 2$，则当$(n,T) \to \infty$. 时， $\sqrt{nT}(\hat{\beta}_{\mathsf{DP}} - \beta) \xrightarrow{d} N(\overline{0,V})$

我们注意到，定理4.6(b)中误差界右侧的第一项是使用 $\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i$ 估计 β 时通常的统计误差率，而第二项是隐私误差，或称“隐私代价”。– 在（5）条件下，无论 n 和 T 的相对增长情况如何，当 $n \to \infty$ 时，这一隐私代价都会变得可以忽略不计。

此外，在定理4.6(c)的结果中，注意到如果所有用户的协变量都具有相同的一阶和二阶矩，即对于公共矩阵 Γ、 Λ， $\Gamma_i = \Gamma$ 和 $\Lambda_i = \Lambda$ ，那么渐近协方差满足 $V = \Gamma^{-1}\Lambda\Gamma^{-1}$。这等于通常的普通最小二乘估计量 β_{OLS} 的渐近协方差，因此，定理4.6表明，与该情形下的非隐私普通最小二乘估计量相比， β_{DP} 在渐近意义下不会损失效率。直接推论是，在替代估计量 β_{SSDP} 的隐私噪声随 T 衰减的情形下，我们的估计量的隐私噪声也会渐近地如此衰减。

4.3. 用户级差分隐私双组回归估计。 除了式 (1) 中介绍的线性回归模型外，另一种值得关注的模型还纳入了时不变协变量。由于每个用户对每个时不变变量仅有一个观测值，因此并不明显能否以用户级隐私的方式估计该变量的效应，并使隐私误差随面板长度 T 增大而衰减。下文中，我们聚焦于这样一个简单情形：两组（例如对照组和处理组）具有不同的截距和/或斜率：

$$y_{i,t} = (\beta + \theta z_i)^{\mathsf{T}} x_{i,t} + \varepsilon_{i,t}, \tag{6}$$

其中， $z_i \in \{0,1\}$ 表示组别归属。我们注意到， θ 原则上可能是稀疏的，因此，如果 $x_{i,t}$ 具有截距项，则模型 (6) 将模型 $y_{i,t} = \beta^{\mathsf{T}} x_{i,t} + \theta z_i + \varepsilon_{i,t}$ （ $\theta \in \mathbb{R}$）作为其特例。为便于分析，我们假设 $n_1 := \sum_{i=1}^{n} z_i$ 和 $n_0 := n - n_1$ 具有相同的量级。

假设 4.7。 我们有 n_0， $n_1 = \Omega(n)$，并且 $\frac{n_1}{n} \to p \in (0,1)$ 当 $n \to \infty$ 时。

如果组别归属变量 z_i 是公开的，那么可以分别对组 1 和组 2 中的用户应用算法 1 来估计 θ ，再对相应的估计值求差。如果组别归属变量 z_i 本身具有敏感性，那么这种方法无法提供隐私保证。算法 4 对算法 1 进行了修改，以便在保护隐私的同时区分具有 $z_i = 1$ 和 $z_i = 0$ 的用户；分别对两组中的用户应用该算法，可得到以下隐私与效用保证。

定理 *4.8.* 设 $\hat{}\beta_i$ 如式 *(2)* 所示，并设 $\hat{}\beta_{1,\mathsf{DP}} = \mathrm{DPTreatMean}((D_i, z_i)_{i=1}^n, \mu/\sqrt{2}, B, R, \xi/4)$ 和 $\hat{}\beta_{0,\mathsf{DP}} = \mathrm{DPTreatMean}((D_i, 1-z_i)_{i=1}^n, \mu/\sqrt{2}, B, R, \xi/4)$。定义 $\hat{}\theta_{\mathsf{DP}} := \hat{}\beta_{1,\mathsf{DP}} - \hat{}\beta_{0,\mathsf{DP}}$。

$\hat{}$***(a)*** 隐私。 θ_{DP} 满足 μ-*uGDP*。***(b)*** 误差界。 在假设 *4.1-4.2*、*4.7* 下，且若

$$B \geq B_{\min} + \|\theta\|, \; R \geq R_{\min}, \; n \geq n_{\min}, \; \text{且} \; T \geq T_{\min},$$

则以至少 $1-\xi$ 的概率，令 $M_j := \frac{\sum_{i:z_i=j} \|\Gamma_i^{-\frac{1}{2}}\|^2 \sum_{i:z_i=j} 1}{}$，

$$\|\hat{\theta}_{\mathsf{DP}} - \theta\| = O\Big(\sigma_\varepsilon\Big[\sqrt{\frac{(M_0+M_1)(d+\log\frac{1}{\xi})}{nT}} + \max_{1\leq i\leq n}\|\Gamma_i^{-\frac{1}{2}}\|\sqrt{\frac{(R\log\frac{R}{\xi}+d)(d+\log\frac{n}{\xi})}{n^2\mu^2T}}\Big]\Big).$$

(c) 渐近正态性。 若进一步假设满足假设 *4.4*，其中 $k=2$，且条件 (5)成立，则若 $V_1 = \lim_{n\to\infty} \frac{1}{p^2 n}\sum_{i:z_i=1}\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}$和 $V_0 = \lim_{n\to\infty} \frac{1}{(1-p)^2 n}\sum_{i:z_i=0}\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}$存在，则有 $\sqrt{nT}(\hat{}\theta_{\mathsf{DP}} - \theta) \xrightarrow{d} N(0, V_1+V_0)$ 当 $(n, T)\to\infty$.

证明见附录 C.4。我们指出，该估计量的隐私误差（即定理 4.8(b) 中误差界的第二项）会随 T 增大而衰减；而在对汇总统计量进行隐私化以构造 β 的私有估计量时，这一点无法实现。为说明这一点，考虑简单情形 $y_{i,t} = \beta^\mathsf{T} x_{i,t} + \theta z_i + \varepsilon_{i,t}$，其中 $\frac{1}{nT}X^\mathsf{T}X = \begin{bmatrix} \frac{1}{nT}\sum_{i,t} x_{i,t}x_{i,t}^\mathsf{T} & \frac{1}{nT}\sum_{i,t} z_i x_{i,t} \\ \frac{1}{nT}\sum_{i,t} z_i x_{i,t}^\mathsf{T} & \frac{1}{n}\sum_{i=1}^n z_i \end{bmatrix}$。显然，只有左上角分块的统计量能够带来所需的一致集中性，从而使隐私误差随 T 增大而衰减。

5. Differentially Private HAC inference

由于可能存在跨期依赖，涉及跨时间观测值的线性回归问题通常需要估计异方差和自相关一致（HAC）协方差矩阵。在纵向回归中，通常通过所谓的聚类稳健协方差矩阵估计量来解决这一问题，该估计量将 $\hat{}\beta_{\mathsf{OLS}}$ 的渐近协方差估计为 $Q^{-1}\hat{}WQ^{-1}$，其中 $Q = \frac{1}{n}\sum_{i=1}^n X_i^\mathsf{T}X_i$ 和 $\hat{}W = \frac{1}{n}\sum_{i=1}^n X_i^\mathsf{T}\hat{}\varepsilon_i\hat{}\varepsilon_i^\mathsf{T}X_i$（Liang 和 Zeger，1986年；Arellano，1987年；Wooldridge，2003年）。Hansen（2007年）证明，在 $(n, T)\to\infty$ 机制下，该估计量是 $\sqrt{n}$-一致的。遗憾的是，这种协方差估计无法直接应用于我们的情形，因为（i）我们的估计量 β_{DP} 具有不同的渐近协方差（见定理 4.6）；（ii）我们无法轻易控制 $\frac{1}{T}X_i^\mathsf{T}X_i$ 和 $\frac{1}{T}X_i^\mathsf{T}\hat{}\varepsilon_i\hat{}\varepsilon_i^\mathsf{T}X_i$ 的敏感度，而这正是私有噪声校准的关键步骤。

因此，我们提出了一种简单的私有估计量，用于估计私有点估计量的渐近协方差；在我们的设定下，该估计量也具有异方差和自相关一致性，并且是 $\sqrt{n}$-一致的。我们不计算三明治型估计量，而是提出一种样本协方差估计量的私有版本

在 $\hat{}\,\beta_i$ 上，即对 $(\hat{}\,\beta_i - \hat{}\,\beta_{\mathsf{DP}})(\hat{}\,\beta_i - \hat{}\,\beta_{\mathsf{DP}})^\mathsf{T} \approx (X_i^\mathsf{T}X_i)^{-1}X_i^\mathsf{T}\varepsilon_i\varepsilon_i^\mathsf{T}X_i(X_i^\mathsf{T}X_i)^{-1}$ 取样本均值。因此，我们的估计量对用户取平均，以规避异方差和自相关问题。这与聚类稳健标准误的思路类似，但 $\hat{}$ 还考虑了 β_i 之间的离散程度。

算法 2 uGDP 协方差矩阵估计

$\hat{}\,\hat{}$ 1: **输入：** 用户级局部估计 β_i、算法 1 的输出空间 $(\beta_{\mathsf{DP}}, \hat{}\,m_{r^*-1}, r^*, B^*)$，以及算法 1 的量 n_{LB} 和初始半径 B，以及协方差隐私参数 μ_{var}。 2: 令 $\delta_{\mathsf{DP}} \leftarrow \|\hat{}\,\beta_{\mathsf{DP}} - \hat{}\,m_{r^*-1}\|$, $\kappa \leftarrow \frac{B}{2^{r^*}} + \delta_{\mathsf{DP}}$, $S \leftarrow \{i \in [n] : \|\hat{}\,\beta_i - \hat{}\,m_{r^*-1}\| \le \frac{B}{2^{r^*}}\}$, $n_{\tilde{S}} \leftarrow \max(|S|, n_{LB})$ 3: 令

$$\tilde{V}(\hat{\beta}_{\mathsf{DP}})_{\mathsf{DP}} \leftarrow \frac{1}{n_{\tilde{S}}^2}\sum_{i\in S}(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} + (B^*)^2 I + W,$$

其中 $W_{i,i} \overset{iid}{\sim} N(0, \frac{16\kappa^4}{n_{LB}^4\mu_{\mathsf{var}}^2})$ 和 $W_{i,j} \overset{iid}{\sim} N(0, \frac{8\kappa^4}{n_{LB}^4\mu_{\mathsf{var}}^2})$，对于 $j > i$，其中 $W_{i,j} = W_{j,i}$ 对于 $j < i$。 4: 令 $\hat{}\,V(\hat{}\,\beta_{\mathsf{DP}})_{\mathsf{DP}}$ 为 $\tilde{V}(\hat{}\,\beta_{\mathsf{DP}})_{\mathsf{DP}}$ 在锥 $\mathbb{M}_{symm}^{d\times d}$ 上的投影。 $\hat{}\,\hat{}$ 5: **输出空间：** $V(\beta_{\mathsf{DP}})_{\mathsf{DP}}$

$\hat{}$ 特别地，我们将算法 2 应用于估计 β_{DP} 的协方差问题。请注意，算法 2 第 3 行中的协方差估计量包含一个小样本校正项 $(B^*)^2 I$，如下一小节所述，其中 B^* 是算法 1 的输出半径。由于算法 2 本质上是对纳入 β_{DP} 估计中的用户局部估计进行的私有样本协方差估计量，因此该算法要求先对 β_{DP} 进行私有估计。

$\hat{}$ $\hat{}\,\hat{}$**定理 5.1.** 设 β_{DP} 采用 μ_{est}-*uGDP* 进行估计。令 $(\beta_i)_{i=1}^n$ 和 β_{DP} 为$\hat{}$ $\hat{}$算法 *2* 的输入，其隐私级别为 μ_{var}，并将所得估计记为 $V(\beta_{\mathsf{DP}})_{\mathsf{DP}}$。

(a) 隐私。 $(\hat{}\,\beta_{\mathsf{DP}}, \hat{}\,V(\hat{}\,\beta_{\mathsf{DP}})_{\mathsf{DP}})$ 满足 $\sqrt{\mu_{\mathsf{est}}^2 + \mu_{\mathsf{var}}^2}$-*uGDP*。***(b)*** 收敛。 若满足定理 *4.6* 的所有条件，则 $(\hat{}\,V(\hat{}\,\beta_{\mathsf{DP}})_{\mathsf{DP}})^{-\frac{1}{2}}(\hat{}\,\beta_{\mathsf{DP}} - \beta) \xrightarrow{d} N(0, I_d)$ 当 $(n,T) \to \infty$ 时成立。 $^-$

定理 4.6 和 5.1 的结合使我们能够对 β 进行渐近有效的推断。由于我们拥有渐近正态的隐私估计量，以及它们的渐近方差的隐私估计量，因此可以得到隐私渐近枢轴量。通过后处理，我们可以基于这些隐私近似枢轴量发布多个置信区间，而不会造成额外的隐私损失。以下推论将采用此方法构造的隐私置信区间的保证形式化。

推论 5.2. 满足定理 *5.1* 的条件。则当 $(n,T) \to \infty$

$$\mathbb{P}\Big((\hat{\beta}_{\mathsf{DP}})_j - z_{1-\alpha/2}\sqrt{(\hat{V}(\hat{\beta}_{\mathsf{DP}})_{\mathsf{DP}})_{j,j}} \le \beta_j \le (\hat{\beta}_{\mathsf{DP}})_j + z_{1-\alpha/2}\sqrt{(\hat{V}(\hat{\beta}_{\mathsf{DP}})_{\mathsf{DP}})_{j,j}}\Big) \xrightarrow{p} 1-\alpha,$$

且这些置信区间对所有 $\alpha \in (0,1)$ 均满足 $\sqrt{\mu_{\text{est}}^2 + \mu_{\text{var}}^2}$-*uGDP*。

此外，人们通常还会关注检验如下形式的假设 $H_0 : R\beta = r$，其中$R \in \mathbb{R}^{q\times d}$是一个秩为 q 的矩阵，q 是在 H_0条件下对 β 施加的线性无关约束的数量。由于我们的估计量渐近正态，且我们对渐近协方差矩阵有一致估计，因此可以将这些结果代入 Wald 检验，从而得到隐私假设检验。以下推论对此进行了形式化表述。

推论 **5.3.** 满足定理 *5.1* 的条件。于是，在 H_0下，当$(n,T) \to \infty$

$$\hat{W}_{\text{DP}} := (R\hat{\beta}_{\text{DP}} - r)^{\mathsf{T}}(R\hat{V}(\hat{\beta}_{\text{DP}})_{\text{DP}}R^{\mathsf{T}})^{-1}(R\hat{\beta}_{\text{DP}} - r) \xrightarrow{d} \chi_q^2,$$

且这些检验对所有 $R \in \mathbb{R}^{q\times d}$ 和 $r \in \mathbb{R}^q$均满足 $\sqrt{\mu_{\text{est}}^2 + \mu_{\text{var}}^2}$-*uGDP*。

5.1. **小样本校正**。 在实践中，考虑一种类似于Avella-Medina 等(2023年提出的有限样本校正方法是有益的。事实上，估计量的“基础部分”是一个在高概率下具有如下形式的估计量

$$\hat{V} := \frac{1}{|S|^2}\sum_{i\in S}(\hat{\beta}_i - \hat{\beta}_{\text{DP}})(\hat{\beta}_i - \hat{\beta}_{\text{DP}})^{\mathsf{T}} = \frac{1}{|S|^2}\sum_{i\in S}(\hat{\beta}_i - \bar{\beta}_S)(\hat{\beta}_i - \bar{\beta}_S)^{\mathsf{T}} + \frac{(B^*)^2}{|S|}ZZ^{\mathsf{T}},$$

其中，S 包含其数据用于计算 β_{DP}的用户集合，即未被截尾的用户。其期望为 $\mathbb{E}[\hat{V}] \approx \text{Var}(\bar{\beta}_S) + \left(\frac{B^*}{|S|}\right)^2 I$。另一方面，以高概率有$\hat{\beta}_{\text{DP}} = \frac{1}{|S|}\sum_{i\in S}\hat{\beta}_i + B^*Z$，其中 $Z \sim N(0, I_d)$，使得$\text{Var}(\hat{\beta}_{\text{DP}}) \approx \text{Var}(\bar{\beta}_S) + (B^*)^2 I$，因此我们得出$\text{Var}(\hat{\beta}_{\text{DP}}) - \mathbb{E}[\hat{V}] \approx (1 - \frac{1}{|S|})(B^*)^2 I \approx (B^*)^2 I$.。因此，我们将该量用作协方差估计过程中的偏差校正项。

5.2. 双组模型中的推断。 稍作修改后，即可扩展该算法及其保证，使其能够对式（6）的双组模型进行推断。具体而言，可以按照第 4.3 节中的方法，分别估计$\hat{\beta}_{0,\text{DP}}$ 和$\beta_{1,\text{DP}}$，并将其分别应用于算法 5（即略作修改的算法 2），从而得到私有协方差估计量$\hat{V}(\beta_{0,\text{DP}})_{\text{DP}}$ 和 $V(\beta_{1,\text{DP}})_{\text{DP}}$。由于用户之间（因而组间）相互独立，$\hat{V}(\theta_{\text{DP}})_{\text{DP}} := V(\beta_{0,\text{DP}})_{\text{DP}} + V(\beta_{1,\text{DP}})_{\text{DP}}$ 在定理 4.8 的条件下给出一致估计。精确的定理表述请参见附录E。

5.3. **讨论**。 在独立同分布误差且协变量二阶矩相等的情况下，可以简单地使用算法 1 对 $\frac{1}{T}X^{\mathsf{T}}X$ 和 σ^2进行隐私估计，从而得到一个 $\sqrt{nT}$-一致的 β_{DP}渐近协方差估计量 $\hat{}$。更有趣的是，利用本文始终采用的工具和机制，或许可以估计 Andrews (1991年) 的 HAC 估计方法的隐私类比形式。我们推测，这可能得到一个 $\sqrt{nT}$-一致估计量

用于协方差矩阵的估计，且需满足适当的正则条件。对这一问题的正式探讨超出了本文的范围，将留待未来研究。

此外，我们指出，假设 4.1 和 4.2 中的尾部假设可以通过多种方式放宽，但代价是有限样本误差界会变弱。具体而言，只要求误差具有截至某一阶数的有限阶矩，因此甚至可以允许多项式尾部。同样，协变量分布的尾部也可以像例如 Wong 等人 (2020年那样，始终假设为次 Weibull；但若不发展新的集中性结果，就无法进一步放宽这一假设，因为我们对 Gram 矩阵的控制依赖于Merlev`ede 等人 (2011) 关于指数强混合次 Weibull 序列的集中性结果。

6. Simulations

6.1. ARMA 估计。我们考虑模型 $y_{i,t} = \beta^\top x_{i,t} + \varepsilon_{i,t}$，其中 $d = 4$，并从 $\beta_j \overset{iid}{\sim}$Unif(−20, 20) 中抽取样本。在此模型中，$(x_{i,t},\varepsilon_{i,t})_{t=1}^{T}$ 在截面上独立同分布，且我们生成$m_i \overset{iid}{\sim} N(0, 9I)$ 和 $x_{i,t} = m_i + 0.5(x_{i,t-1} - m_i) + \eta_{i,t}$，其中 $\eta_{i,t} \overset{iid}{\sim} N(0, I)$。最后，误差遵循模型 $\varepsilon_{i,t} = 0.5\varepsilon_{i,t-1} + 0.5r_{i,t-1} + r_{i,t}$，其中 $r_{i,t} \overset{iid}{\sim} N(0, 1)$。我们考察此模型中 β 的各种估计量的性能，即：定理 4.6 中的差分隐私估计量 β_{DP}，其隐私参数为 $\mu = 1$-uGDP；隐私参数为 $\mu = \infty$-uGDP 的差分隐私估计量 β_{DP}（即无隐私）；以及普通最小二乘估计量 $\beta_{\mathsf{OLS}} = (X^\top X)^{-1}X^\top Y$。表 1 展示了这些估计量在 1000 次独立重复试验中的经验缩放均方根误差$\sqrt{nT}\|\hat{}\beta - \beta\|$，图 3 则展示了这些重复试验中的经验未缩放均方根误差，即 $\|\beta - \beta\|$。在算法 1 的所有实例中，我们使用了 $B = 100, R = 10$ 和 $\xi = 10^{-5}$。

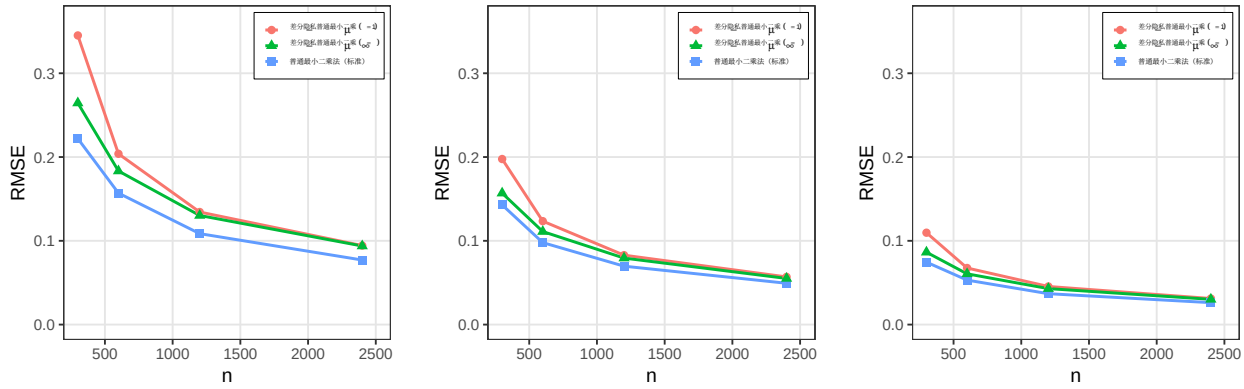
Table 1. β的各种估计量对应的$\sqrt{nT}\|\hat{}\beta - \beta\|$经验值。

	差分隐私普通最小二乘法（$\mu = 1$）			差分隐私普通最小二乘法（$\mu = \infty$）			普通最小二乘法（$\mu = \infty$）		
$n\backslash T$	10	40	160	10	40	160	10	40	160
300	18.91	21.65	24.03	14.49	17.19	18.93	12.16	15.65	16.33
600	15.79	19.13	20.93	14.20	17.20	18.77	12.16	15.16	16.45
1200	14.72	18.15	19.94	14.27	17.39	18.81	11.89	15.27	16.16
2400	14.61	17.63	19.30	14.53	17.06	18.66	11.94	15.28	16.22

回想一下，根据 ˆ定理 4.6，β_{DP} 满足以下误差界，其中 T 足够大 ：

$$\sqrt{nT}\|\hat{\beta}_{\mathsf{DP}} - \beta\| \lesssim C_{\mathsf{stat}} + \frac{C_{\mathsf{priv}}}{\sqrt{n}},$$

其中， C_{stat} 表示使用 $\frac{1}{n}\sum_{i=1}^{n}\hat{}\beta_i$估计 β 时通常的标准化统计误差，而 $\frac{1}{\sqrt{n}}C_{\text{priv}}$ 表示标准化隐私代价。具体而言，隐私代价项 $\frac{1}{\sqrt{n}}C_{\text{priv}}$ 反映了在定理 4.6 中采用隐私级别 $\mu = 1$ 相较于采用$\mu = \infty$ 所产生的额外误差。这与我们在表 1 和图 3 中的发现一致：我们 ^ ^ 观察到，随着 n增大，无论 T取何值，采用 $\mu = 1$ 的 β_{DP} 的均方根误差都会收敛到采用 $\mu = \infty$ 的 β_{DP} 的均方根误差。此外，我们观察到，在所有方法中——无论是否采用隐私保护——增加 n 带来的收益都大于增加 T带来的收益。或许，这是因为增加 T 会引入新的相关数据，而增加 n 则会引入一个新的独立数据序列；在有限样本下，后者更有益。



(A) $T = 10$ (B) $T = 40$ (C) $T = 160$

FIGURE 3. 各种（n， T）组合的未缩放均方根误差。

6.2. ARMA 推断。我们采用上一小节的模型，但现在固定 $T = 15$，同时也考虑 $n = 4800$，并使用算法 2 中描述的聚类协方差估计量，评估置信区间 $(\beta_j - z_{0.975}V_{j,j}, \beta_j + z_{0.975}V_{j,j})$ 的经验覆盖率和宽度；根据推论 5.2，这些区间渐近覆盖率为 95%。我们评估基于定理 5.1 的区间，这些区间满足 $\sqrt{2}$-uGDP（取 $\mu_{\text{est}} = \mu_{\text{var}} = 1$）以及 ∞- uGDP（即完全不采用隐私保护）。和之前一样，我们在估计 $\hat{}\beta_{\text{DP}}$ 时取 $B = 100$、 $R = 10$ 和 $\xi = 10^{-5}$。我们将这些结果与非隐私合并普通最小二乘法估计量 $\beta_{\text{OLS}} = (X^\top X)^{-1}X^\top Y$ 达到的覆盖率进行比较：该估计量采用非异方差和自相关一致协方差估计 $\hat{}\sigma^2(X^\top X)^{-1}$，我们称之为“经典”方法；此外，还比较了基于聚类稳健标准误（例如 Arellano（1987年））的方法，我们称之为“聚类”方法。表 2 列出了这四组估计量达到的覆盖率，以及构造的置信区间的平均宽度（括号内）。

该表基于每种配置进行 10,000 次重复模拟，因此在 $d = 4$ 下，每种配置共有 40,000 个区间。对于 $n = 4800$，隐私区间的覆盖率达到 95%，其宽度几乎与非隐私对应方法相同，也只比使用合并普通最小二乘法的聚类稳健标准误构造的置信区间略宽。

TABLE 2. ARMA 95% 置信区间的经验覆盖率（及宽度）。注：覆盖率下方括号内报告的是宽度。

	差分隐私普通最小二乘法		合并普通最小二乘法	
n	$\sqrt{2}$-uGDP	∞-uGDP	聚类	经典方法
300	0.895 (0.524)	0.951 (0.456)	0.947 (0.413)	0.526 (0.151)
600	0.936 (0.375)	0.949 (0.323)	0.948 (0.293)	0.525 (0.107)
1200	0.947 (0.246)	0.949 (0.228)	0.951 (0.207)	0.524 (0.075)
2400	0.949 (0.167)	0.950 (0.162)	0.950 (0.147)	0.525 (0.053)
4800	0.950 (0.116)	0.950 (0.114)	0.950 (0.104)	0.520 (0.038)

7. DATA EXAMPLE

我们考虑的数据示例是儿童哮喘管理计划（CAMP，2000）数据集的匿名教学版本。该数据集包含 $n = 695$ 名哮喘儿童在多次医生就诊期间的肺活量测量值。在此期间，所有儿童均接受了阿苯诺治疗。此外，约有 2/5 未接受其他药物治疗，其余 3/5 接受了某种其他实验性药物治疗。主要测量指标是第一秒用力呼气量（FEV）。我们感兴趣的问题是，在接受处理组治疗后，该指标如何随时间变化。因此，我们考虑的第一个回归模型是，

$$\mathrm{FEV}_{i,t} = \alpha + \alpha_i + \beta \mathrm{days}_{i,t} + \varepsilon_{i,t}, \tag{7}$$

其中，用力呼气量$_{i,\ t}$ 表示儿童在第 i次就诊时的用力呼气量，t^{th} 表示从儿童开始接受处理组治疗到其$_{i,\ t}$ 次就诊时所经过的天数。我们关注的主要统计量是 $_i$。我们使用两种估计量对其进行估计。第一种是 t^{th} 非私有单向随机效应估计量，假设 β的均值为零且独立同分布，并且，该模型使用 R 中的 plm 软件包拟合（Croissant

和 Millo，2008年），去均值参数通过 Swamy 和 Arora（1972年）的方法获得。对于该估计量，我们使用 Arellano（1987年）中的聚类稳健ˆ协方差估计量计算标准误。第二种估计量是 β_{DP} ，如定理 4.6 所述，使用 $\mu = 1$、 $B = 1$、 $R = 15$和 $\xi = 10^{-6}$进行估计。我们使用定理 5.1 的估计量，并结合 $\mu_{\mathsf{var}} = 1$ 来获得标准误，因此（ˆβ_{DP}，ˆV(ˆβ_{DP})$_{\mathsf{DP}}$）满足 $\sqrt{2}$-uGDP。表 3 列出了用这两种方法得到的点估计和标准误。由于私有估计量必然具有随机性，差分隐私条目包含 1000 次重复实验所得估计值的四分位距。非私有以及全部 1000 个私有 z 分数均强烈拒绝 $H_0 : \beta = 0$。另外请注意，这是一个包含患者的非平衡面板

TABLE 3. （7）的估计回归系数

估计量	估计系数	SE
$\hat{\beta}_{\mathsf{RE}}$	0.000640	$8.687 \cdot 10^{-6}$
$\hat{\beta}_{\mathsf{DP}}$	$[0.000639,\ 0.000645]$	$[8.192 \cdot 10^{-6},\ 8.638 \cdot 10^{-6}]$

随访次数介于 0 次至 17 次之间，其中 8 名用户的数据点少于 3 个。明确考虑这种不平衡面板可能会提高效率。

我们继续探讨本研究的首要问题：额外的药物治疗是否会对用力呼气量产生影响。在这种情况下，相关的回归方程为

$$\mathsf{FEV}_{i,t} = \alpha + z_i\rho + \alpha_i + (\beta + z_i\theta)\mathrm{days}_{i,t} + \varepsilon_{i,t} =: \varphi^{\mathsf{T}} x_{i,t} + \varepsilon_{i,t}, \tag{8}$$

其中，如果患者 i 接受了额外的药物治疗，则 z_i 等于 1；如果患者 i 属于对照组，则等于 0。我们关注的量是 θ，它衡量接受额外药物治疗的患者斜率系数的增幅。我们的策略是使用算法 4，对治疗组中的用户斜率进行隐私估计（对应于 $\beta+\theta$），对对照组中的用户也进行同样的估计，然后取两者之差，以获得 θ的估计值。我们同样运行两次算法 5，并将结果相加，以获得方差估计。由于我们使用 $\frac{1}{\sqrt{2}}$-uGDP 分别运行了两次算法 4 和算法 5，因此结果最终的隐私保证为 $\sqrt{2}$-uGDP。此外，我们的分析基于小样本：安慰剂组有 $n_1 = 275$，非安慰剂组有$n_0 = 420$，因此我们采用较低的 $R = 10$。分析结果见表 4。我们注意到， θ_{RE} 在 $\alpha = 0.05$ 水平上不具有统计学显著性，其 p 值为 0.248；而私有的 z 检验在 1000 次重复实验中有 65 次拒绝了 $H_0 : \theta = 0$ 。

TABLE 4. 式（8）的斜率差系数估计值

估计量	估计系数	SE
$\hat{\theta}_{\mathsf{RE}}$	$-2.043 \cdot 10^{-5}$	$1.770 \cdot 10^{-5}$
$\hat{\theta}_{\mathsf{DP}}$	$[-2.634 \cdot 10^{-6},\ 5.119 \cdot 10^{-5}]$	$[4.095 \cdot 10^{-5},\ 4.453 \cdot 10^{-5}]$

汇总统计量的隐私化。估计 (8) 的另一种方法是分别计算 $X^\mathsf{T}X$ 和 $X^\mathsf{T}Y$的用户级差分隐私估计，然后将它们组合成一个线性回归估计量。具体而言，对于每位用户，令

$$x_{i,t,1} = \mathrm{days}_{i,t} - \overline{\mathrm{days}}_i, \qquad x_{i,t,2} = z_i x_{i,t,1}, \qquad \text{and} \qquad y_{i,t} = \mathsf{FEV}_{i,t} - \overline{\mathsf{FEV}}_i,$$

从而得到固定效应估计量的差分隐私对应形式。我们使用算法 1 估计 $(X^\mathsf{T}X)_{\mathsf{DP}}$ 和 $(X^\mathsf{T}Y)_{\mathsf{DP}}$，利用 $X_i^\mathsf{T}X_i$ 的上三角部分作为 $\mathbb{R}^3$中的输入，并将差分隐私输出转换为 $\mathbb{R}^{2\times2}$中的矩阵。对于这两个量，我们在算法 1 中将初始半径 B 设为满足关系 $\max_{i=1,\ldots,695} \|D_i\| \le B$ 的最小可能 B，该关系适用于我们的数据集；并设置 $R = 1$（对于 R的各种取值，由于任何变量都不存在围绕共同量的集中性，因此半径最多收缩一次），以及 $\mu = 1$ 和 $\xi = 10^{-5}$，以得到最终估计 $\hat{\varphi}_{\mathsf{SSOptimistic}} := (X^\mathsf{T}X)^{-1}_{\mathsf{DP}}(X^\mathsf{T}Y)_{\mathsf{DP}}$。这个估计量并非真正满足差分隐私，因为$B$ 和 R 的校准是根据观测数据中的最小可能值进行的，而非根据全局量进行的。1000 次重复实验的估计结果见表5。

TABLE 5. 使用 $\phi_{\mathsf{SSOptimistic}}$ 对 (8) 进行估计所得的系数

系数 估计系数 四分位距	
$\hat{\beta}_{\mathsf{SSOptimistic}}$	$[0.000636,\ 0.000672]$
$\hat{\theta}_{\mathsf{SSOptimistic}}$	$[-8.041 \cdot 10^{-5},\ -4.869 \cdot 10^{-6}]$

表5显示，即使是通过向汇总统计量添加噪声构造的乐观估计量，其估计结果也比基于汇总局部平均值的隐私估计量噪声更大；后者的估计值列于表3和表4。我们指出，在该数据集中，T 较小；正如第4.2节所讨论的，基于汇总局部平均值的隐私估计量具有一项优势：其隐私误差会随 T 衰减。因此，我们预计随着面板长度增加，或面板变得更加平衡，β_{DP} 与 $\beta_{\mathsf{SSOptimistic}}$ 相比会表现得更好。

8. Discussion

目前已有大量文献探讨稳健统计与差分隐私之间的联系，以开发新的隐私算法。例如（Dwork 和 Lei，2009；Avella-Medina 和 Brunel，2020；Avella-Medina，2020，2021；Liu 等，2022；Ramsay 等，2022；Avella-Medina 等，2023；Li 等，2023；Hopkins 等，2023；Alabi 等，2023；Chhor 和 Sentenac，2023；Yu 等，2024），这里只列举其中一些。我们指出，尽管我们的核心算法基于 Karwa 和 Vadhan（2018）以及 Biswas 等人（2020）的差分隐私思想，但其构造直观上蕴含一定程度的稳健性，并且与 Gervini 和 Yohai（2002）提出的稳健且高效估计思想有些相似。若能对我们方法的稳健性进行正式刻画，将会很有意义。

此外，我们还讨论了线性回归估计量的效率。已有多项研究（Dobriban 和 Sheng，2021年；Rosenblatt 和 Nadler，2016年）探讨了分布式环境中平均估计量的效率。Dobriban 和 Sheng (2021年)指出，尽管在我们的设定中，当 $\Gamma_i = \Gamma\ \forall i = 1, \ldots\ldots, n$时，渐近效率不会损失，但存在协变量异质性时则并非如此。此外，他们的结果表明，通常情况下，简单平均估计量（采用与我们研究相同的等权重）在渐近意义上不如最优加权的平均线性回归估计量高效。由于每个最优权重都是特定于某个用户的参数，要以差分隐私方式实施此类加权方案并非易事，但沿此方向开展研究或可得到效率更高的隐私估计量。

此外，尽管我们的工作研究了用户级差分隐私（uDP）下线性回归问题的高效估计与推断，将其推广到广义线性模型和 M 估计量也是很自然的。这些基础模型在更简单的项目级差分隐私设定下已受到一定关注（Lei，2011；Cai 等，2020；Avella-Medina，2020、2021；Arora 等，2022；Avella-Medina 等，2023），许多研究也探讨了差分隐私经验风险最小化和随机凸优化中的预测任务（Chaudhuri 等，2011；Kifer 等，2012；Bassily 等，2014、2019）。近期论文也在用户级差分隐私设定下研究了其中一些问题（Levy 等，2021；Bassily 和 Sun，2023；Ghazi 等，2023；Liu 和 Asi，2024），但未涉及推断或相关数据。

类似地，非隐私纵向回归文献中已有大量研究，无法轻易纳入我们的隐私设定。这些研究包括检验序列相关是否存在及其形式（Baltagi 和 Li，1995年；Hong 和 Kao，2004年；Baltagi 等，2007年）、截面相关（Frees，1995年；Ng，2006年；Moscone 和 Tosetti，2009年；Baltagi 等，2016年）、斜率同质性（Pesaran 和 Yamagata，2008年），以及单位根是否存在（Maddala 和 Wu，1999年；Levin 等，2002年；Pesaran，2007年；Bai 和 Ng，2010年），

以及许多其他相关程序。研究这些方法的差分隐私对应方法，将具有重要的实际意义。

最后，在误差协方差矩阵 Σ 可以得到一致估计的情形下，标准回归理论表明，本文提出的估计量可通过纳入可行广义最小二乘机制得到渐近改进。我们在这一方向上取得了初步且令人鼓舞的成果，相关内容将另文发表。

References

Abowd, J. M. (2018年). 美国人口普查局采用差分隐私. 载于知识发现与数据挖掘国际会议（*KDD*），第2867–2867页。Agarwal, S., Kamath, G., Majid, M., Mouzakis, A., Silver, R., and Ullman, J. (2025年). 个人级差分隐私下的私有均值估计. 载于离散算法研讨会（*SODA*）。Alabi, D., Kothari, P. K., Tankala, P., Venkat, P., and Zhang, F. (2023年). 高斯分布的私有估计：高效、稳健且最优. 载于计算理论研讨会（*STOC*）。Alabi, D., McMillan, A., Sarathy, J., Smith, A., and Vadhan, S. (2022年). 差分隐私简单线性回归. 隐私增强技术期刊（*PoPETs*）。Alabi, D. G. and Vadhan, S. P. (2023年). 线性回归的差分隐私假设检验.《机器学习研究杂志》，24(361):1–50。Amin, K., Joseph, M., Ribero, M., and Vassilvitskii, S. (2023年). 简易差分隐私线性回归. 载于国际学习表征会议（*ICLR*）。Andrews, D. W. (1991年). 异方差和自相关稳健协方差矩阵估计. *Econometrica*，第817–858页。Arellano, M. (1987年). 组内估计量稳健标准误的计算.《牛津经济学与统计学通报》，49(4)。Arora, R., Bassily, R., Guzm´an, C., Menart, M., and Ullah, E. (2022年). 重新审视差分隐私广义线性模型. 神经信息处理系统进展（*NeurIPS*），第22505–22517页。Autor, D. H., Dorn, D., Hanson, G. H., and Song, J. (2014年). 贸易调整：工人层面证据.《经济学季刊》，129(4):1799–1860。Avella-Medina, M. (2020年). 稳健统计在隐私数据分析中的作用.《*Chance*》，33(4):37–42。Avella-Medina, M. (2021). 隐私保护参数推断：稳健统计的用武之地.《美国统计协会会刊》，116(534):969–983。Avella-Medina, M., Bradshaw, C., and Loh, P.-L. (2023年). 通过噪声优化实现差分隐私推断.《统计年鉴》，51(5):2067–2092。Avella-Medina, M. and Brunel, V.-E. (2020年). 提出、检验、发布：具有高概率保证的差分隐私估计. *arXiv*预印本 *arXiv:2002.08774*。

Bai, J. and Ng, S. (2010年). 存在截面依赖的面板单位根检验：进一步研究. 《计量经济理论》, 26(4):1088–1114。Baltagi, B. H., Kao, C., and Peng, B. (2016年). 存在序列相关的大型面板数据模型中的截面相关检验. 《计量经济学》, 4(4):44。Baltagi, B. H. and Li, Q. (1995年). 在误差分量模型中检验 AR(1) 扰动项与 MA(1) 扰动项. 《计量经济学杂志》, 68(1):133–151。Baltagi, B. H., Song, S. H., Jung, B. C., and Koh, W. (2007年). 使用面板数据检验序列相关、空间自相关和随机效应. 《计量经济学杂志》, 140(1):5–51。Barrientos, A. F., Reiter, J. P., Machanavajjhala, A., and Chen, Y. (2019年). 回归系数的差分隐私显著性检验. 《计算与图形统计学杂志》, 28(2):440–453。Barrientos, A. F., Williams, A. R., Snoke, J., and Bowen, C. M. (2024年). 差分隐私汇总统计量和回归分析的可行性研究：基于行政数据和调查数据的评估. 《美国统计协会会刊》, 119(545):52–65。Bassily, R., Feldman, V., Talwar, K., and Guha Thakurta, A. (2019年). 具有最优速率的私有随机凸优化. 神经信息处理系统进展（*NeurIPS*）。Bassily, R., Smith, A., and Thakurta, A. (2014年). 私有经验风险最小化：高效算法与紧确误差界. 载于计算机科学基础研讨会（*FOCS*）, 第464–473页。Bassily, R. and Sun, Z. (2023年). 具有最优速率的用户级私有随机凸优化. 载于国际机器学习会议, 第1838–1851页。PMLR。Biswas, S., Dong, Y., Kamath, G., and Ullman, J. (2020年). CoinPress：实用的私有均值与协方差估计. 载于神经信息处理系统进展（*NeurIPS*）。Brockwell, P. J. and Davis, R. A. (1991年). 《时间序列：理论与方法》。施普林格科学与商业媒体。Cai, T. T., Wang, Y., and Zhang, L. (2020年). 广义线性模型中的隐私代价：算法与极小极大下界. *arXiv*预印本 *arXiv:2011.03900*。Cai, T. T., Wang, Y., and Zhang, L. (2021年). 隐私代价：差分隐私下参数估计的最优收敛速率. 《统计年刊》, 49(5):2825–2850。Cai, Z., Li, S., Xia, X., and Zhang, L. (2023年). 高维回归中的私有估计与推断及错误发现率控制. *arXiv*预印本 *arXiv:2310.16260*。CAMP (2000). 布地奈德或奈多罗米对哮喘儿童的长期影响. 《新英格兰医学杂志》, 343(15):1054–1063。Chaudhuri, K., Monteleoni, C., and Sarwate, A. D. (2011). 差分隐私经验风险最小化. 《机器学习研究杂志》, 12(3)。

Chhor, J. and Sentenac, F.（2023）。局部差分隐私下离散分布的稳健估计。载于算法学习理论国际会议（*ALT*），第411–446页。Church, T. S., Blair, S. N., Cocreham, S., Johannsen, N., Johnson, W., Kramer, K., Mikus, C. R., Myers, V., Nauta, M., Rodarte, R. Q., et al.（2010）。有氧运动和抗阻训练对2型糖尿病患者糖化血红蛋白水平的影响：一项随机对照试验。*JAMA*，304(20):2253–2262。Cormode, G., Jha, S., Kulkarni, T., Li, N., Srivastava, D., and Wang, T.（2018）。大规模隐私保护：实践中的局部差分隐私。载于数据管理国际会议，第1655–1658页。Croissant, Y. and Millo, G.（2008）。R 中的面板数据计量经济学：plm软件包。《统计软件杂志》，27:1–43。Cummings, R., Feldman, V., McMillan, A., and Talwar, K.（2022）。数据异质性下的用户级隐私均值估计。载于神经信息处理系统进展（*NeurIPS*）。Dobriban, E. and Sheng, Y.（2021）。通过平均实现分布式线性回归。《统计年鉴》，49(2):918–943。Dong, J., Roth, A., and Su, W. J.（2022）。高斯差分隐私。《皇家统计学会杂志》*B*辑，84(1):3–37。Doukhan, P.（1995）。《混合：性质与示例》。施普林格。Dwork, C.（2006）。差分隐私。载于自动机、语言与编程国际研讨会（*ICALP*）。Dwork, C. and Lei, J.（2009）。差分隐私与稳健统计。载于计算理论研讨会（*STOC*）。Dwork, C. and Roth, A.（2014）。差分隐私的算法基础。基础与趋势® 理论计算机科学，9(3–4):211–407。Fan, J., Qi, L., and Tong, X.（2016）。弱相关数据下的惩罚最小二乘估计。《中国科学：数学》，59(12):2335–2354。Flynn, K. E., Pi˜na, I. L., Whellan, D. J., Lin, L., Blumenthal, J. A., Ellis, S. J., Fine, L. J., Howlett, J. G., Keteyian, S. J., Kitzman, D. W., et al.（2009）。运动训练对慢性心力衰竭患者健康状况的影响：HF-ACTION 随机对照试验。*JAMA*，301(14):1451–1459。Frees, E. W.（1995）。评估面板数据中的截面相关。《计量经济学杂志》，69(2):393–414。Gervini, D. and Yohai, V. J.（2002）。一类稳健且完全有效的回归估计量。《统计年鉴》，30(2):583–616。Ghazi, B., Kamath, P., Kumar, R., Manurangsi, P., Meka, R., and Zhang, C.（2023）。每位用户只有少量样本时的用户级差分隐私。神经信息处理系统进展（*NeurIPS*）。

Hansen, C. B.（2007）。当 t 较大时面板数据稳健方差矩阵估计量的渐近性质。《计量经济学杂志》，141(2):597–620。Hong, Y. and Kao, C.（2004）。基于小波的面板模型未知形式序列相关检验。*Econometrica*，72(5):1519–1563。Hopkins, S. B., Kamath, G., Majid, M., and Narayanan, S.（2023）。稳健性意味着统计估计中的隐私保护。载于计算理论研讨会（*STOC*）。Hsu, D., Kakade, S. M., and Zhang, T.（2012）。岭回归的随机设计分析。载于学习理论会议（*COLT*）。Jacobson, L. S., LaLonde, R. J., and Sullivan, D. G.（1993）。被裁员工的收入损失。《美国经济评论》，第685–709页。Kamath, G., Li, J., Singhal, V., and Ullman, J.（2019）。私有高维分布学习。载于学习理论会议（*COLT*），第1853–1902页。Karwa, V. and Vadhan, S.（2018）。有限样本差分隐私置信区间。载于理论计算机科学创新会议（*ITCS*）。Kent, A., Berrett, T. B., and Yu, Y.（2024）。用户级局部差分隐私的速率最优性与相变。*arXiv*预印本 *arXiv:2405.11923*。Kifer, D., Smith, A., and Thakurta, A.（2012）。私有凸经验风险最小化与高维回归。载于学习理论会议，第25–1页。JMLR Workshop and Conference Proceedings。Lei, J.（2011）。差分隐私 M 估计量。神经信息处理系统进展（*NeurIPS*）。Levin, A., Lin, C.-F., and Chu, C.-S. J.（2002）。面板数据中的单位根检验：渐近性质与有限样本性质。《计量经济学杂志》，108(1):1–24。Levy, D., Sun, Z., Amin, K., Kale, S., Kulesza, A., Mohri, M., and Suresh, A. T.（2021）。用户级隐私保护学习。载于神经信息处理系统进展（*NeurIPS*）。Li, M., Berrett, T. B., and Yu, Y.（2023）。关于稳健性与局部差分隐私。《统计年鉴》，51(2):717–737。Liang, K.-Y. and Zeger, S. L.（1986）。使用广义线性模型进行纵向数据分析。《生物统计学》，73(1):13–22。Liu, D. and Asi, H.（2024）。用户级差分隐私随机凸优化：具有最优速率的高效算法。载于人工智能与统计国际会议（*AISTATS*）。Liu, W., Zhang, Y., Yang, H., and Meng, Q.（2024）。医疗数据分析中的差分隐私综述。《数据科学年鉴》，11(2):733–747。Liu, X., Kong, W., and Oh, S.（2022）。高维情形下的差分隐私与稳健统计。载于学习理论会议（*COLT*）。Maddala, G. S. and Wu, S.（1999）。面板数据单位根检验的比较研究及一种新的简易检验。《牛津经济学与统计学公报》，61(S1):631–652。

Merlev`ede, F., Peligrad, M., and Rio, E.（2011）。弱依赖序列的伯恩斯坦型不等式与中等偏差。《概率论及相关领域》，151(3):435–474。Mewes, J., Fairbrother, M., Giordano, G. N., Wu, C., and Wilkes, R.（2021）。经历很重要：关于美国个人层面社会信任下降原因的纵向研究。《社会科学研究》，95:102537。Milionis, J., Kalavasis, A., Fotakis, D., and Ioannidis, S.（2022）。无界协变量下的差分隐私回归。载于人工智能与统计国际会议（*AISTATS*）。Mokkadem, A.（1988）。ARMA过程的混合性质。《随机过程及其应用》，29(2):309–315。Moscone, F. and Tosetti, E.（2009）。面板中截面独立性检验的综述与比较。《经济调查杂志》，23(3):528–561。Ng, S.（2006）。利用间距统计量检验面板数据中的截面相关性。《商业与经济统计杂志》，24(1):12–23。Olver, F. W.（2010）。*NIST*数学函数手册（精装版及*CD-ROM*）。剑桥大学出版社。Pesaran, M. H.（2007）。存在截面依赖时的一种简单面板单位根检验。《应用计量经济学杂志》，22(2):265–312。Pesaran, M. H. and Yamagata, T.（2008）。检验大规模面板中的斜率同质性。《计量经济学杂志》，142(1):50–93。Ramsay, K., Jagannath, A., and Chenouri, S.（2022）。差分隐私多元中位数。*arXiv*预印本 *arXiv:2210.06459*。Rosenblatt, J. D. and Nadler, B.（2016）。论分布式统计学习中平均方法的最优性。信息与推断：*IMA*期刊，5(4):379–404。Roth, V. and Avella-Medina, M.（2025）。依赖数据下的差分隐私。*arXiv*预印本，*arXiv:2511.18583*。Sheffet, O.（2019）。差分隐私线性回归中的传统技术。载于算法学习理论（*ALT*）。Sobel, M. E.（2012）。婚姻会提高男性工资吗？：面板数据固定效应回归模型中处理效应的识别。《美国统计协会会刊》，107(498):521–529。Swamy, P. A. V. B. and Arora, S. S.（1972）。误差成分回归模型中系数估计量的精确有限样本性质。*Econometrica*：《计量经济学学会杂志》，第261–275页。Vershynin, R.（2018）。《高维概率：数据科学导论及其应用》，第47卷。剑桥大学出版社。von Bahr, B. and Esseen, C.-G.（1965）。随机变量之和的 r阶绝对矩不等式，$1 \le r \le 2$。《统计数学年鉴》，第299–303页。

Wainwright, M. J.（2019）。《高维统计：一种非渐近视角》，第48卷。剑桥大学出版社。Wang, Y.-X.（2018）。重新审视差分隐私线性回归：无界域中的最优与自适应预测和估计。*arXiv*预印本 *arXiv:1803.02596*。Wasserman, L. and Zhou, S.（2010）。差分隐私的统计框架。《美国统计协会会刊》，105(489):375–389。Wong, K. C., Li, Z., and Tewari, A.（2020）。β-混合重尾时间序列的Lasso保证。《统计年鉴》，48(2):1124–1142。Wooldridge, J. M.（2003）。应用计量经济学中的整群抽样方法。《美国经济评论》，93(2):133–138。Yu, M., Ren, Z., and Zhou, W.-X.（2024）。高斯差分隐私稳健均值估计与推断。《伯努利》，30(4):3059–3088。

APPENDIXA. AUXILIARY RESULTS

在 A.1 节中，我们首先回顾一些著名的集中不等式和几个线性代数引理，它们将在主要结果的证明中发挥作用。A.2 节介绍强混合条件下的集中不等式，这些不等式构成了支撑我们理论分析的主要技术工具。

A.1. 基本不等式与集中性。

引理 A.1. （*Wainwright, 2019*, 命题 *2.5*）设 X_i 对于 $i \in [n]$ 是相互独立的 σ_i^2-次高斯随机变量，其均值为 m_i。那么，对于所有 $q \geq 0$,

$$\mathbb{P}\Big(\sum_{i=1}^{n}(X_i - m_i) \geq q\Big) \leq \exp\Big(-\frac{q^2}{2\sum_{i=1}^{n}\sigma_i^2}\Big) \quad \text{and} \quad \mathbb{P}\Big(\sum_{i=1}^{n}(X_i - m_i) \leq -q\Big) \leq \exp\Big(-\frac{q^2}{2\sum_{i=1}^{n}\sigma_i^2}\Big).$$

引理 A.2. （*Wainwright, 2019*, 例 *2.1* 和式（*2.9*））设 $Z_r \overset{iid}{\sim} subG$（1）满足 $\mathbb{E}[Z_r] = 0$对于 $r \in [R]$。那么，对于任意 $0 < \xi \leq 1$,

$$\mathbb{P}\Big(\max_{r\in[R]}|Z_r| \leq \sqrt{2\log\frac{R}{\xi}}\Big) \geq 1 - 2\xi.$$

引理 A.3。 *(Wainwright, 2019年, 例 2.28)* 设 $N_r \overset{iid}{\sim} N(0,\ \mathbb{I}_d)$ 其中 $r \in [R]$。那么，对于任意 $0 < \xi \leq 1$,

$$\mathbb{P}\Big(\max_{r\in[R]}\|N_r\| \leq \sqrt{d} + \sqrt{2\log\frac{R}{\xi}}\Big) \geq 1 - \xi.$$

引理 A.4。 定义函数 $g : [0,1] \mapsto \mathbb{R}_+$。设 $Z \in \mathbb{R}^d$ 是一个随机向量，使得对于所有 $u \in \mathbb{S}^{d-1}$ 和 $\xi \in [0,1]$，都有 $\mathbb{P}(|u^\intercal Z| > g(\xi)) \leq \xi$，则

$$\mathbb{P}\Big(\|Z\| > \frac{4}{3}g\Big(\frac{\xi}{9^d}\Big)\Big) \leq \xi.$$

同样，若 $W \in \mathbb{R}^{d\times d}$ 是一个对称的半正定随机矩阵，使得对于所有 $u \in \mathbb{S}^{d-1}$ 和 $\xi \in [0,1]$，都有 $\mathbb{P}(|u^\intercal W u| > g(\xi)) \leq \xi$，则

$$\mathbb{P}\Big(\|W\| > 2g\Big(\frac{\xi}{9^d}\Big)\Big) \leq \xi.$$

证明。 这可由标准的 ϵ-网论证得到。具体而言，根据 Vershynin（2018年）的推论 4.2.13，存在一个集合 $\mathbf{N}$，使得 $\sup_{x\in\mathbb{S}^{d-1},\ y\in\mathcal{N}}\|x-y\| \leq \frac{1}{4}$ 且 $|\mathbf{N}| \leq 9^d$。现在，取$v = Z/\|Z\|$和 $v_0 \in \mathbf{N}$，使得 $\|v - v_0\| \leq \frac{1}{4}$。那么 $|v_0^\intercal Z| \geq \sup_{v\in\mathbb{S}^{d-1}}|v^\intercal Z| - |(v-v_0)^\intercal Z| \geq (1-\frac{1}{4})\|Z\|$，因此 $\sup_{y\in\mathcal{N}}|y^\intercal Z| \geq \frac{3}{4}\|Z\|$。于是，

$$\mathbb{P}\Big(\|Z\| > \frac{4}{3}g(\xi)\Big) \leq \mathbb{P}\Big(\sup_{y\in\mathcal{N}}|y^\intercal Z| > g(\xi)\Big) \leq |\mathcal{N}|\,\xi,$$

因此，通过取 $\xi' = \xi/9^d$，我们得到

$$\mathbb{P}\Big(\|Z\| > \frac{4}{3}g(\xi')\Big) = \mathbb{P}\Big(\|Z\| > \frac{4}{3}g\Big(\frac{\xi}{9^d}\Big)\Big) \leq |\mathcal{N}|\,\frac{\xi}{9^d} \leq \xi.$$

矩阵情形下的证明类似，注意到 $\|W\| = \sup_{u\in\mathbb{S}^{d-1}} u^\top W u$，并且对于两个单位向量 u，u_0，满足 $\|u-u_0\| \le \frac{1}{4}$，我们有 $u^\top W u = u_0^\top W u_0 + (u-u_0)^\top W u_0 + u^\top W (u-u_0)$，因此 $|u^\top W u - u_0^\top W u_0| \le \frac{1}{2}\|W\|$。□

引理 A.5。 *(Vershynin, 2018年，练习 7.3.5)* 设 $W \in \mathbb{R}^{d\times d}$ 是一个随机矩阵，使得 $W_{i,i} \overset{iid}{\sim} N(0, 2\sigma^2)$ 对所有 $i \in [d]$ 都成立，并且 $W_{i,j} \overset{iid}{\sim} N(0, \sigma^2)$ 对所有 $i > j$ 都成立，其中 $W_{i,j} = W_{j,i}$，当 $i < j$ 时。则 $\|W\| = O(\sigma\sqrt{d + \log\frac{1}{\xi}})$ 以至少 $1-\xi$ 的概率成立。

引理 A.6。 设 A 和 B 是实对称可逆矩阵，满足 $\|B^{-\frac{1}{2}} A B^{-\frac{1}{2}} - I\| \le \epsilon < 1$. 那么，我们有 $\|B^{\frac{1}{2}} A^{-\frac{1}{2}}\| \le (1-\epsilon)^{-1/2}$.

证明。 注意到 $\|B^{\frac{1}{2}} A^{-\frac{1}{2}}\| = \sqrt{\|B^{\frac{1}{2}} A^{-\frac{1}{2}}\|^2} = \sqrt{\|B^{\frac{1}{2}} A^{-1} B^{\frac{1}{2}}\|}$ 此外，根据假设，我们有 $(1-\epsilon) I \le B^{-\frac{1}{2}} A B^{-\frac{1}{2}} \le (1+\epsilon) I$，因此取逆并相应地反转不等号方向，可得

$$\frac{1}{1-\epsilon} I \ge B^{\frac{1}{2}} A^{-1} B^{\frac{1}{2}} \ge \frac{1}{1+\epsilon} I.$$

因此，我们有

$$\|B^{\frac{1}{2}} A^{-\frac{1}{2}}\| = \sqrt{\|B^{\frac{1}{2}} A^{-1} B^{\frac{1}{2}}\|} \le \sqrt{\frac{1}{1-\epsilon}} = \frac{1}{\sqrt{1-\epsilon}}.$$

□

引理 A.7。 设 $Z \in \mathbb{R}$ 为零均值的 σ^2-次高斯变量。则 $\mathbb{E}[\text{指数}(\frac{Z^2}{6\sigma^2})] \le 2$。

证明。 设 $a = \frac{1}{6\sigma^2}$。则，

$$\begin{aligned}\mathbb{E}[\exp(aZ^2)] &= \int_0^\infty \mathbb{P}(\exp(aZ^2) > u) du = 1 + \int_1^\infty \mathbb{P}(\exp(aZ^2) > u) du \\ &= 1 + \int_0^\infty 2at e^{at^2} \mathbb{P}(|Z| > t) dt \le 1 + 4a \int_0^\infty t e^{-(\frac{1}{2\sigma^2} - a)t^2} dt = 1 + \frac{2a}{\frac{1}{2\sigma^2} - a},\end{aligned}$$

通过变量替换 $u = e^{at^2}$ 并应用引理 A.1 的尾界。由于我们采用双侧尾界，引理 A.1 尾界右侧还需乘以额外因子 2。现在代入 $a = \frac{1}{6\sigma^2}$，即可得到结果 $\mathbb{E}[\text{指数}(\frac{Z^2}{6\sigma^2})] \le 2$. □

引理 A.8。 设随机变量 X 和 $Y \in \mathbb{R}$ 分别为 σ_X^2 和 σ_Y^2-次高斯，且对于某个常数 $c > 0$，有 $|\mathbb{E}[X]| \le c\sigma_X$ 和 $|\mathbb{E}[Y]| \le c\sigma_Y$，则 $\mathbb{P}(|XY - \mathbb{E}[XY]| > q) \le 2\exp(-\frac{q}{C\sigma_X\sigma_Y})$，其中常数 $C > 0$ 仅取决于 c。

证明。 令 $m_X := \mathbb{E}[X]$、$m_Y := \mathbb{E}[Y]$、$\tilde{X} = X - m_X$、$\tilde{Y} = Y - m_Y$。则有 $XY - \mathbb{E}[XY] = A + B + C$、$\tilde{A} = \tilde{X}\tilde{Y} - \mathbb{E}[\tilde{X}\tilde{Y}]$、$\tilde{B} := m_X \tilde{Y}$、$\tilde{C} = m_Y \tilde{X}$。我们首先证明 $m_X, m_Y \neq 0$ 的情形。显然可得，

$$\mathbb{P}(|A| > q/3) \le 2e^{-\frac{q}{36\sigma_X\sigma_Y}}, \tag{9}$$

$$\mathbb{P}(|B| > q/3) \le 2e^{-\frac{q}{18|m_X|\sigma_Y}}, \tag{10}$$

$$\mathbb{P}(|C| > q/3) \le 2e^{-\frac{q}{18|m_Y|\sigma_X}}. \tag{11}$$

事实上，(10)–(11) 直接由引理 A.1 得出。因此，我们证明 (9)。对于任意 $u, v \in \mathbb{R}$、$2|uv| \le u^2 + v^2$，有

$$\Big|\frac{\tilde{X}\tilde{Y}}{\sigma_X\sigma_Y}\Big| \le \frac{1}{2}\Big(\frac{\tilde{X}^2}{\sigma_X^2} + \frac{\tilde{Y}^2}{\sigma_Y^2}\Big).$$

取期望并应用柯西-施瓦茨不等式，利用引理 A.7，可得：

$$\mathbb{E}\Big[\exp\Big(\frac{|\tilde{X}\tilde{Y}|}{6\sigma_X\sigma_Y}\Big)\Big] \le 2,$$

因此，$\mathbb{P}(|\tilde{X}\tilde{Y}| > q) \le 2\exp(-\frac{q}{6\sigma_X\sigma_Y})$ 对所有 $q > 0$ 均成立。此外，根据柯西-施瓦茨不等式，$|\mathbb{E}[\tilde{X}\tilde{Y}]| \le \sigma_X\sigma_Y$. 现在，首先考虑 $q > 2|\mathbb{E}[\tilde{X}\tilde{Y}]|$。于是 $q - |\mathbb{E}[\tilde{X}\tilde{Y}]| \ge q/2$，并且

$$\mathbb{P}(|\tilde{X}\tilde{Y} - \mathbb{E}[\tilde{X}\tilde{Y}]| > q) \le \mathbb{P}(|\tilde{X}\tilde{Y}| > q/2) \le 2\exp(-\frac{q}{12\sigma_X\sigma_Y}).$$

现在令 $0 < q \le 2|\mathbb{E}[\tilde{X}\tilde{Y}]|$。显然有

$$2\exp\Big(-\frac{q}{12\sigma_X\sigma_Y}\Big) \ge 2\exp\Big(-\frac{2|\mathbb{E}[\tilde{X}\tilde{Y}]|}{12\sigma_X\sigma_Y}\Big) \ge 2\exp\Big(-\frac{2\sigma_X\sigma_Y}{12\sigma_X\sigma_Y}\Big) = 2e^{-1/6} \ge 1;$$

因此，由于每个概率显然也都被 1 所界定，我们得出，对于所有 $q > 0$,

$$\mathbb{P}(|\tilde{X}\tilde{Y} - \mathbb{E}\tilde{X}\tilde{Y}| > q) \le 2\exp(-\frac{q}{12\sigma_X\sigma_Y}).$$

通过取 $q/3$，由此证明了式（9）。现在，对式（9）—（11）中的事件应用并集界，

$$\begin{aligned}\mathbb{P}(|XY - \mathbb{E}[XY]| > q) &\le 2e^{-\frac{q}{36\sigma_X\sigma_Y}} + 2e^{-\frac{q}{18|m_X|\sigma_Y}} + 2e^{-\frac{q}{18|m_Y|\sigma_X}} \le 2e^{-\frac{q}{36\sigma_X\sigma_Y}} + 4e^{-\frac{q}{18c\sigma_X\sigma_Y}} \\ &\le 6e^{-\frac{q}{c'\sigma_X\sigma_Y}},\end{aligned}$$

其中 $c' = \max(36, 18c)$。

我们需要得到形式为 $\mathbb{P}(|XY - \mathbb{E}[XY]| > q) \le 2\exp(-\frac{q}{C\sigma_X\sigma_Y})$，的尾界。根据我们的推导，并且由于概率不超过 1，我们因此需要证明 $\min(1, 6\exp(-\frac{q}{c'\sigma_X\sigma_Y})) \le 2\exp(-\frac{q}{C\sigma_X\sigma_Y})$。当 $C = c'\log 6 \log 2$ 时，该不等式成立，从而证明了给定 $C > 0$ 下的结论。

注意，如果 m_X、m_Y 中有一个或两个为零，那么相应的概率界（关于 B 或 C，或两者）显然为零，因此该界在这种情况下也成立。□

引理 A.9。 *(von Bahr 和 Esseen，1965年，定理 2）* 设 $Z_1, \ldots, Z_n$ 为一列满足 $\mathbb{E}[Z_{m+1} \mid \sum_{i=1}^m Z_m] = 0$ 几乎处处对所有 $1 \le m \le n-1$成立的随机变量，且 $\mathbb{E}[|Z_i|^r] < \infty$对所有 $i \in [n]$ 以及某个 $1 \le r \le 2$成立。则，$\mathbb{E}[|\sum_{i=1}^n Z_i|^r] \le 2\sum_{i=1}^n \mathbb{E}[|Z_i|^r]$.

A.2. **强混合与集中性。**

引理 A.10. 若 $(z_t)_{t\in\mathbb{Z}}$ 是强混合的，且其强混合系数为 $\alpha(k)$，并且 $g : \mathbb{R}^d \mapsto \mathbb{R}^p$ 是可测函数，则 $(g(z_t))_{t\in\mathbb{Z}}$ 也具有强混合性，其强混合系数为 $\alpha_g(k) \le \alpha(k)$。

证明。 根据定义 2.10，我们有

$$\alpha_g(k) = \sup_{t\in\mathbb{Z}} \sup_{A\in\mathcal{G}_{-\infty}^{t}, B\in\mathcal{G}_{t+k}^{\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

其中，$\mathsf{G}_a^b = \sigma(g(X_a)\ldots, g(X_b))$ 是由 $g(X_a)$ 经由 $g(X_b)$ 生成的 σ-代数。显然，我们有 $\mathsf{G}_a^b \subseteq \mathsf{F}_{a}^{b}$,其中 F_a^b 是相应的、由X_a 经由 X_b生成的 σ-代数。因此，根据定义，我们得到

$$\begin{aligned}\alpha_g(k) &= \sup_{t\in\mathbb{Z}}\Big(\sup_{A\in\mathcal{G}_1^t, B\in\mathcal{G}_{t+k}^{\infty}} \big|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\big|\Big)\\ &\le \sup_{t\in\mathbb{Z}}\Big(\sup_{A\in\mathcal{F}_1^t, B\in\mathcal{F}_{t+k}^{\infty}} \big|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\big|\Big) = \alpha(k)\end{aligned}$$

□

引理 A.11。 *(Doukhan，1995年，第 1.2 章，定理 3；Davydov 不等式）*设$(X_t)_{t\in\mathbb{Z}}$是一个随机过程。则，假设 $\mathbb{E}[\|X_s\|^p] < \infty$ 且 $\mathbb{E}[\|X_t\|^q] < \infty$ 对于 $s,\ t \in \mathbb{Z}$,

$$|\mathrm{Cov}(X_s, X_t)| \le 8\alpha(|s-t|)^{\frac{1}{r}}(\mathbb{E}[\|X_s\|^p])^{1/p}(\mathbb{E}[\|X_t\|^q])^{1/q},$$

对于任意满足 $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} = 1$ 的 $p, q, r \ge 1$，且其中 $\alpha(\cdot)$ 是该过程$(X_t)_{t\in\mathbb{Z}}$ 的 α-混合系数，如定义 *2.10* 中所定义。

引理 *A.12*。（*Doukhan，1995*年，第 *1.4* 章，定理 *2*）设 $\tau > 2$，且 $\mathcal{C}$ 是 $\mathbb{N}_+$ 的有限子集。此外，设 $(Z_t)_{t\in\mathbb{Z}}$ 是一列中心化强混合随机变量，其强混合系数为 $\alpha(k)$，满足 $\exists\epsilon > 0$, $\exists c \in 2\mathbb{N}$，其中 $c \ge \tau$，且 $\sum_{r=1}^{\infty}(r+1)^{c-2}\alpha(r)^{\epsilon/(c+\epsilon)} < \infty$，并且 $\max_{t\in\mathcal{C}}\mathbb{E}[|Z_t|^{\tau+\epsilon}] < \infty$。那么，

$$\mathbb{E}|\sum_{t\in\mathcal{C}} Z_t|^{\tau} \le C\max\Big(\sum_{t\in\mathcal{C}}(\mathbb{E}|Z_t|^{\tau+\epsilon})^{\tau/(\tau+\epsilon)}, \Big(\sum_{t\in\mathcal{C}}(\mathbb{E}|Z_t|^{2+\epsilon})^{2/(2+\epsilon)}\Big)^{\tau/2}\Big) = O(|\mathcal{C}|^{\tau/2})$$

引理 *A.13*。（*Merlev`ede* 等，*2011*，定理 *1*）设 $(Z_t)_{t\in\mathbb{Z}}$ 是一列实值中心化随机变量，满足对某些常数 γ_2、$b > 0$ 以及所有 $q > 0$，有 $\sup_{t>0}\mathbb{P}(|Z_t| > q) \le \exp\{1-(q/b)^{\gamma_2}\}$；此外，其强混合系数 $\alpha(k)$ 满足 $\alpha(k) \le \exp\{-ck^{\gamma_1}\}$，其中 c、$\gamma_1 > 0$ 为某些常数。那么，记

$$V := \sup_{M>0}\sup_{t>0}\Big(\mathrm{Var}\,(\tau_M(Z_t)) + 2\sum_{r>t}\big|Cov(\tau_M(Z_t), \tau_M(Z_r))\big|\Big),$$

令 $\tau_M(x) := \max\{-M, \min(x, M)\}$，对于由 $1/\gamma = 1/\gamma_1 + 1/\gamma_2$ 定义的 $T \ge 4$ 和 $\gamma < 1$，以及依赖于 c, b, γ_1, γ_2 的正常数 C_1, C_2, C_3, C_4，对所有 $q > 0$,

$$\mathbb{P}\Big(\sup_{R\le T}|\sum_{t=1}^{R} Z_t| > q\Big) \le T\exp\Big\{-\frac{q^{\gamma}}{C_1}\Big\} + \exp\Big\{-\frac{q^2}{C_2(1+TV)}\Big\} + \exp\Big\{-\frac{q^2}{C_3 T}\Big(\frac{q^{\gamma(1-\gamma)}}{C_4(\log q)^{\gamma}}\Big)\Big\}.$$

引理 A.14。 设$(Z_t)_{t\in\mathbb{Z}}$ 是一列中心化强混合随机变量，其强混合系数为 $\alpha(k) \le C\exp\{-ck\}$。此外，若对于某个 $K > 0$ 及所有 $q > 0$,

我们有 $\sup_t \mathbb{P}(|Z_t| > q) \le 2e^{-\frac{q}{K}}$，那么对于任意 $T > 0$，存在一个不依赖于 T 或 q的常数 $c_K > 0$，使得

$$\mathbb{P}\Big(\Big|\frac{1}{T}\sum_{t=1}^{T} Z_t\Big| > q\Big) \le T\exp\Big(\frac{-(Tq)^{\frac{1}{2}}}{c_K}\Big) + \exp\Big(\frac{-(Tq)^2}{c_K(1+TK^2)}\Big) + \exp\Big(\frac{-Tq^2}{c_K}\exp\Big(\frac{(Tq)^{\frac{1}{4}}}{c_K(\log(Tq))^{\frac{1}{2}}}\Big)\Big).$$

特别地，以至少 $1-\xi$ 的概率，对于依赖于 K 的常数 C_K，

$$\Big|\frac{1}{T}\sum_{t=1}^{T} Z_t\Big| = O\Big(K\sqrt{\frac{1}{T}\log\frac{1}{\xi}} + \frac{C_K}{T}\Big(\log\frac{T}{\xi}\Big)^2\Big).$$

证明。 注意到根据假设$\gamma_1 = 1$ 且 $\gamma_2 = 1$，因此 $\gamma = \frac{1}{2}$。实际上， $\gamma_2 = 1$ 由引理 A.13 可直接推出：

$$\sup_{t>0}\mathbb{P}(|Z_t| > q) \le 2e^{-\frac{q}{K}} \le ee^{-\frac{q}{K}} = \exp\Big\{1 - \frac{q}{K}\Big\}.$$

最后，为了界定其量 V，对于截断参数 $\tau_M(x) = \max\{\min\{x, M\}, -M\}$，

$$\begin{aligned} V &:= \sup_{M>0}\sup_{t>0}\Big(\mathrm{Var}(\tau_M(Z_t)) + 2\sum_{r>t}|\mathrm{Cov}(\tau_M(Z_t), \tau_M(Z_r))|\Big) \\ &\le \sup_{t>0}\Big(\mathrm{Var}(Z_t) + 2\sum_{r>t}|\mathrm{Cov}(Z_t, Z_r)|\Big) \le \sup_{t>0}\Big(\mathbb{E}[Z_t^2] + 2C\mathbb{E}[Z_t^2]\sum_{k\ge1}\alpha(k)^{\frac{1}{2}}\Big) = O(K^2), \end{aligned}$$

应用次指数随机变量的矩界，并使用引理 A.11（Davydov 不等式）得到最终不等式，再利用指数强混合假设可得 $\sum_{k\ge1}\alpha(k)^{\frac{1}{2}} = O(1)$。□

引理A.15。 设 $x_{i,t}$ 对于 $i \ge 1, t \in [T]$ 如假设 ***4.1*** 所述，且$(\varepsilon_{i,t})$ 如假设 ***4.2*** 所述。那么，记作 $\Gamma_{i,T} = \frac{1}{T}X_i^\top X_i$，对于 $T \ge 3$ 和 $2 \le p \prec \infty$，我们有上确界$_{i\ge1}\mathbb{E}[\|\Gamma_{i,T} - \Gamma_i\|^p] = O(T^{-p/2})$，以及上确界$_{i\ge1}\mathbb{E}[\|S_{i,T}\|^p] = O(1)$ 记作 $S_{i,T} = \frac{1}{\sqrt{T}}X_i^\top\varepsilon_i$。

证明。 首先注意到 $\mathbb{E}[\|\Gamma_{i,T} - \Gamma_i\|^p] = \int_0^1 Q_i(\xi)^p d\xi$，其中 $Q_i(\xi) := \inf\{x \in \mathbb{R} : F_i(x) \ge \xi\}$ 是分位数函数，如果 $F_i(x) = \mathbb{P}(\|\Gamma_{i,T} - \Gamma_i\| \le x)$。因此，我们首先求出以下对象的尾概率

$$\Big\|\frac{1}{T}\sum_{t=1}^{T}(x_{i,t}x_{i,t}^\top - \mathbb{E}[x_{i,t}x_{i,t}^\top])\Big\|.$$

注意到 $|\mathbb{E}[\frac{u^\top x_{i,t}}{\sqrt{u^\top\Gamma_i u}}]| \le 1$，并且由于 $x_{i,t}$ 是 σ_x^2-次高斯，且 $\lambda_{\max}(\Gamma_i) \le C_\Gamma < \infty$，我们可以应用引理 A.8 得出

$$\mathbb{P}(|u^\top x_{i,t}x_{i,t}^\top u - \mathbb{E}[u^\top x_{i,t}x_{i,t}^\top u]| > q) = \mathbb{P}\Big(\Big|\frac{u^\top x_{i,t}x_{i,t}u - \mathbb{E}[u^\top x_{i,t}x_{i,t}^\top u]}{u^\top\Gamma_i u}\Big| > \frac{q}{u^\top\Gamma_i u}\Big) \le 2\exp\Big(\frac{-q}{C\sigma_x^2 C_\Gamma}\Big).$$

由于根据引理 A.10，$(u^\top x_{i,\,t}x_{i',\,t}^\top u)_{t\in\mathbb{Z}}$ 是指数强混合的，应用引理 A.14 可得，以至少 $1-\xi$ 的概率，对于任意 $u\in\mathbb{S}^{d-1}$，

$$w_h(u) := \Big|\frac{1}{T}\sum_{t=1}^{T} u^\top x_{i,t}x_{i,t}^\top u - u^\top \mathbb{E}[x_{i,t}x_{i,t}^\top]u\Big| = O\Big(\sigma_x^2 C_\Gamma\sqrt{\frac{1}{T}\log\frac{1}{\xi}} + \frac{C_{\sigma_x,\Gamma}}{T}(\log\frac{T}{\xi})^2\Big).$$

由于 $\|\frac{1}{T}\sum_{t=1}^{T}(x_{i,\,t}x_{i',\,t}^\top - \Gamma_i)\| = \sup_{u\in\mathbb{S}^{d-1}} w_h(u)$，根据引理 A.4，我们以至少 $1-\xi$ 的概率得到：

$$\|\Gamma_{i,T}-\Gamma_i\| \le C\Big(\sigma_x^2 C_\Gamma\sqrt{\frac{1}{T}\Big(d+\log\frac{1}{\xi}\Big)} + C_{\sigma_x,\Gamma}\frac{(d+\log\frac{T}{\xi})^2}{T}\Big) =: h(\xi), \tag{12}$$

对于某个 $C>0$。因此，由于 $Q_i(1-\xi)\le h(\xi)$，我们得到

$$\begin{aligned}\mathbb{E}[\|\Gamma_{i,T}-\Gamma_i\|^p] &\le \int_0^1 h(\xi)^p d\xi \\ &\le 2^{p-1}C^p\Big(\int_0^1\Big(\sigma_x^2 C_\Gamma\sqrt{\frac{1}{T}\Big(d+\log\frac{1}{\xi}\Big)}\Big)^p d\xi + \int_0^1\Big(\frac{C_{\sigma_x,\Gamma}}{T}(d+\log\frac{T}{\xi})^2\Big)^p d\xi\Big),\end{aligned}$$

对于第一个积分，令 $\psi=\log\frac{1}{\xi}$ 进行代换，可得

$$\int_0^1\Big(\sigma_x^2\sqrt{\frac{1}{T}\Big(d+\log\frac{1}{\xi}\Big)}\Big)^p d\xi = \sigma_x^{2p}T^{-p/2}\int_0^\infty (d+\psi)^{p/2}e^{-\psi}d\psi = \sigma_x^{2p}T^{-p/2}e^d\Gamma\Big(1+\frac{p}{2},d\Big),$$

其中，$\Gamma(s,\ r) := \int_r^\infty t^{s-1}e^{-t}dt$ 表示不完全伽马函数。因此，第一个项为 $O(T^{-p/2})$；对于第二个项，同样令 $\psi=\log\frac{1}{\xi}$ 进行代换，可得

$$\begin{aligned}\int_0^1\Big(C_{\sigma_x,\Gamma}\frac{(d+\log\frac{T}{\xi})^2}{T}\Big)^p d\xi &= C_{\sigma_x,\Gamma}^p T^{-p}\int_0^\infty (d+\log T+\psi)^{2p}e^{-\psi}d\psi \\ &= C_{\sigma_x,\Gamma}^p T^{-p}e^{d+\log T}\Gamma(2p+1,d+\log T) \le C_{\sigma_x,\Gamma}^p T^{-p+1}e^d\Gamma(2p+1,d),\end{aligned}$$

由于不完全伽马函数关于其第二个参数单调递减，因此对于 $T\ge 3$，有 $\Gamma(2p+1,\ d+\log T)\le\Gamma(2p+1,\ d)$。因此，由于这些项均不依赖于 i，对于 $p\ge 2$，我们有

$$\sup_{i\ge 1}\mathbb{E}[\|\Gamma_{i,T}-\Gamma_i\|^p] = O(T^{-p/2}). \tag{13}$$

继续考察得分 $S_{i,\,T}$ 的矩，首先注意到 $\mathbb{E}[\|S_{i,T}\|^p] = \mathbb{E}\,[\mathbb{E}[\|S_{i,T}\|^p \mid X_i]]$。我们还注意到，对于满足 $\frac{1}{r}+\frac{1}{r'}=1$ 的 $p>2$、$r=p/2$ 和 $r'=\frac{p}{p-2}$，利用 H¨older 不等式将 L_2 范数转为 L_p 范数，可得

$$\|v\|^p = \Big(\sum_{j=1}^d |v_j|^2\cdot 1\Big)^{p/2} \le \Big(\Big(\sum_{j=1}^d |v_j|^{2r}\Big)^{1/r}\Big(\sum_{j=1}^d 1\Big)^{1/r'}\Big)^{p/2} = d^{p/2-1}\sum_{j=1}^d |v_j|^p, \tag{14}$$

现在，在条件 X_i 下，我们有 $S_{i,T}$ 是 $\|\Gamma_{i,T}\|^{\frac{1}{2}}\sigma_\varepsilon$-次高斯，因此，令 $(e_j)_{j=1}^d$ 表示 $\mathbb{R}^d$ 的自然基，并利用 (14)，可得

$$\begin{aligned}\mathbb{E}[\|S_{i,T}\|^p \mid X_i] &\le d^{\frac{p}{2}-1}\sum_{j=1}^d \mathbb{E}[|e_j^\top S_{i,T}|^p \mid X_i] \\ &\le d^{\frac{p}{2}} C_p \sigma_\varepsilon^p \|\Gamma_{i,T}\|^{p/2} p^{p/2} \le d^{\frac{p}{2}} C_p \sigma_\varepsilon^p \left(\|\Gamma_i\| + \|\Gamma_{i,T}-\Gamma_i\|\right)^{\frac{p}{2}} p^{p/2} \\ &\le d^{\frac{p}{2}} C_p 2^{\frac{p}{2}-1}\sigma_\varepsilon^p p^{\frac{p}{2}} \left(\|\Gamma_i\|^{\frac{p}{2}} + \|\Gamma_{i,T}-\Gamma_i\|^{\frac{p}{2}}\right),\end{aligned}$$

因此，对于 d， $p = O(1)$，利用 (13) 以及 $\|\Gamma_i\| \le C_\Gamma$ 在 i上一致这一事实，

$$\sup_{i\ge 1}\mathbb{E}[\|S_{i,T}\|^p] \le d^{\frac{p}{2}} C_p 2^{\frac{p}{2}-1}\sigma_\varepsilon^p p^{\frac{p}{2}}\Big(\sup_{i\ge 1}\|\Gamma_i\|^{\frac{p}{2}} + \sup_{i\ge 1}\mathbb{E}[\|\Gamma_{i,T}-\Gamma_i\|^{\frac{p}{2}}]\Big) = O(1).$$

□

Appendix B. Proofs of Section 3

在本附录中，我们为算法 3 给出隐私保证。算法 3 是算法 1 的轻微推广。具体而言，取 $\mu_1 = \mu_2 = \frac{\mu}{2\sqrt{R}}$ 和 $\mu_3 = \frac{\mu}{\sqrt{2}}$即可从算法 3 得到算法 1。我们还利用该算法推导定理 3.1 的效用保证。

命题 B.1. 算法 *3* 的组合输出空间 $(\hat{m}, \hat{m}_0, \ldots, \hat{m}_{r^*-1}, \hat{m}_{r^*}\mathbf{1}(r^* \ne R)\ r^*)$ 满足 $\sqrt{R\mu_1^2 + R\mu_2^2 + \mu_3^2}$-高斯差分隐私。

证明。 我们先考虑算法在 $r = R$处终止的情况。组合输出仅通过统计量 $\sum_{i=1}^n \omega_i$ 和 $\frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-1})$依赖用户数据。输出最多使用 $2R+1$ 个此类统计量；特别是，如果算法在 $r = R$处终止，则恰好使用数据 $2R+1$ 次。均值细化（第 20 行）最多使用 R 次，用户数量检查（第 7 行）最多使用 R 次，最终输出（第 10 行或第 15 行）使用一次。根据高斯机制的性质，我们只需证明：如果算法在 $r = R$ 处终止，则第 7 行统计量的敏感度为 1，第 10 行和第 15 行统计量的敏感度分别为 $\frac{2B}{2^{r-1}n_{LB}}$ 和 $\frac{2B}{2^R L}$，最后，第 20 行统计量的敏感度为 $\frac{2B}{2^r L}$。

$\sum_{i=1}^n \omega_i$ 的敏感度很简单。 ω_i 是指示变量，每个用户对应一个；更改某个用户的整行数据，最多只能使该指示变量改变 1。因此，该统计量的敏感度为 1，结果是对第 7 行的每次调用都满足 μ_2-高斯差分隐私。这里我们利用了这样一个事实：$\hat{m}_{r-1}$ 已在前一次迭代中单独进行了隐私处理，因此根据后处理，不会泄露更多隐私。

最后，第 10、15 和 20 行都涉及$\frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-j})$的敏感度，其中 S 是 $[n]$ 上的索引集，且 $n_{\tilde{S}}$ =最大为{$|S|$， n_{LB}}，而 $j \in$为{1、2}；在第 10 行中为 $j = 2$，在第 15 和 20 行中为 $j = 1$ 。同样，$\hat{m}_{r-j}$ 已单独进行了隐私处理，因此根据后处理，我们可以将其视为公开量。

不失一般性，假设用户1的数据发生变化。需要考虑四种情况。第一种情况是， $1 \notin S$ 在第一个用户的数据发生变化之前和之后。

算法 3 基于差分隐私修剪的均值估计

输入： 数据 $\{D_i\}_{i=1}^n$，隐私参数 (μ_1, μ_2, μ_3)，半径 B，最大轮数 R，失败概率 ξ

设定： $\tau \leftarrow n - \frac{1}{\mu_2}\sqrt{2\log\frac{4R}{\xi}}$ 和 $n_{LB} \leftarrow$ 最大值 $(2\tau - n, 1)$ ▷ 最终估计中用户数量的高概率下界

最终估计中的用户数

初始化： $\hat{m}_{r-2} \leftarrow 0, \hat{m}_{r-1} \leftarrow 0, r \leftarrow 0, r^* \leftarrow 0$

当 $r \le R$ 且 $r^* = 0$ **时**

▷缩小半径，并检查是否有足够的用户落入新的球内

$\omega_i \leftarrow \mathbb{1}\{\|D_i - \hat{m}_{r-1}\| \le B/2^r\}$，$\omega \leftarrow \sum_{i=1}^n \omega_i + \frac{1}{\mu_2} Z_r$，$Z_r \sim \mathrm{N}(0, 1)$

如果 $\omega < \tau$ **则**

▷过多用户落在球外：输出前一次迭代的半径

$S \leftarrow \{i : \|D_i - \hat{m}_{r-2}\| < B/2^{r-1}\}$，$n_{\tilde{S}} \leftarrow$ 最大值$(|S|, n_{LB})$，$\mathcal{C} \leftarrow \frac{\sqrt{(R-r)\mu_1^2 + (R-r)\mu_2^2 + \mu_3^2}}{\mu_3}$

$\hat{m} \leftarrow \hat{m}_{r-2} + \frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-2}) + \frac{2B}{2^{r-1}\mathcal{C}\mu_3 n_{LB}} N_*$，$N_* \sim \mathrm{N}(0, I)$

$r^* \leftarrow r - 1$

否则如果 $r = R$ **则**

▷最后一轮：输出最小可能半径

$S \leftarrow \{i : \|D_i - \hat{m}_{R-1}\| < B/2^R\}$，$n_{\tilde{S}} \leftarrow$ 最大值$(|S|, n_{LB})$

$\hat{m} \leftarrow \hat{m}_{R-1} + \frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{R-1}) + \frac{2B}{2^R \mu_3 n_{LB}} N_*$，$N_* \sim \mathrm{N}(0, I)$

$r^* \leftarrow R$

else

▷否则：更新中心并缩小半径

$S \leftarrow \{i : \|D_i - \hat{m}_{r-1}\| < B/2^r\}$，$n_{\tilde{S}} \leftarrow$ 最大值$(|S|, n_{LB})$

$\hat{m}_r \leftarrow \hat{m}_{r-1} + \frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-1}) + \frac{2B}{2^r \mu_1 n_{LB}} N_r$，$N_r \sim \mathrm{N}(0, I)$

$r \leftarrow r + 1$

结束 if

结束 while

输出： $(\hat{m}、\hat{m}_0、\ldots\ldots、\hat{m}_{r^*-1}、\hat{m}_{r^*}\mathbf{1}(r^* \ne R)\ \ r^*)$

在这种情况下，敏感度为零。第二种情况是，$1 \in S$ 在第一个用户的数据发生变化之前和之后。在这种情况下，$n_{\tilde{S}} = n'_{\tilde{S}}$，且敏感度有如下界

$$\begin{aligned}\Big\|\frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-j}) - \frac{1}{n'_{\tilde{S}}}\sum_{i\in S'}(D'_i - \hat{m}_{r-j})\Big\| &\le \frac{1}{n_{LB}}\Big\|D_i - \hat{m}_{r-j} - (D'_i - \hat{m}_{r-j})\Big\| \\ &\le \frac{1}{n_{LB}}\Big(\frac{B}{2^{r-(j-1)}} + \frac{B}{2^{r-(j-1)}}\Big) = \frac{2B}{2^{r-(j-1)} n_{LB}}.\end{aligned}$$

第三种情况是，$1 \in S$ 在初始状态下成立，但更改第一个用户的数据后则不成立。在这种情况下，$n'_{\tilde{S}} \le n_{\tilde{S}}$。由于 $n'_{\tilde{S}}$ 的下界为 n_{LB}，因此在这方面最坏情况下的变化是

$n_{\tilde{S}} = n_{LB} + 1$ 和 $n'_{\tilde{S}} = n_{LB}$。那么，敏感度为

$$\begin{aligned}\left\|\frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-j}) - \frac{1}{n'_{\tilde{S}}}\sum_{i\in S'}(D_i - \hat{m}_{r-j})\right\| &= \left\|\left(\frac{1}{n_{\tilde{S}}} - \frac{1}{n'_{\tilde{S}}}\right)\sum_{i\in S'}(D_i - \hat{m}_{r-j}) + \frac{1}{n_{\tilde{S}}}(D_1 - \hat{m}_{r-j})\right\| \\ &\le \left\|\frac{-1}{n_{\tilde{S}}n'_{\tilde{S}}}\sum_{i\in S'}(D_i - \hat{m}_{r-j})\right\| + \left\|\frac{1}{n'_{\tilde{S}}}(D_1 - \hat{m}_{r-j})\right\| \\ &\le \frac{2B}{2^{r-(j-1)}n_{LB}}.\end{aligned}$$

最后，第四种情况是 $1 \notin S$，但 $1 \in S'$。因此，$n'_{\tilde{S}} \ge n_{\tilde{S}} \ge n_{LB}$。同样，最坏情况下的变化发生在 $n_{\tilde{S}} = n_{LB}, n'_{\tilde{S}} = n_{LB} + 1$，因此我们得到

$$\left\|\frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-j}) - \frac{1}{n'_{\tilde{S}}}\sum_{i\in S'}(D_i - \hat{m}_{r-j})\right\| \le \frac{2B}{2^{r-(j-1)}n_{LB}},$$

因为这与前一种情况完全对称。因此，第 10 行中的统计量（其中 $j = 2$）的敏感度上界为 $\frac{2B}{2^{r-(2-1)}n_{LB}} = \frac{2B}{2^{r-1}n_{LB}}$，而第 15 行和第 20 行中的统计量的敏感度上界为 $\frac{2B}{2^{r-(1-1)}n_{LB}} = \frac{2B}{2^r n_{LB}}$。

因此，根据高斯机制，第 7 行的每个实例都满足 μ_2-高斯差分隐私，第 20 行的每个实例都满足 μ_1-高斯差分隐私，而最终输出（第 10 行或第 15 行）满足 μ_3-高斯差分隐私。因此，所有 $\hat{m}_r$、私有计数检查和最终输出的组合满足 μ-高斯差分隐私。由于我们的最终输出（$\hat{m}$, r^*）仅通过这些隐私化统计量依赖于数据，因此根据组合定理，（$\hat{m}$, r^*）满足 $\sqrt{R\mu_1^2 + R\mu_2^2 + \mu_3^2}$-高斯差分隐私。

上述分析涵盖了算法通过达到 $r = R$ 而终止的情况。另一种可能是算法通过进入第 7 行的 if 语句而终止。在这种情况下，为剩余迭代的细化和用户计数检查所预留的隐私预算仍未使用，因此可以重新收回，用于最终估计。具体而言，假设算法通过第 7 行的 if 语句并以 r^* 终止。在这种情况下，算法执行了 $r^* + 1$ 次中心细化，即在第 0 次迭代至第 r^* 次迭代中执行。此外，算法执行了 $r^* + 1$ 次用户计数检查，即在第 1 次迭代至第 r^*+1 次迭代中执行。因此，如果使用隐私参数为 $\mathsf{C}\mu_3$ 的高斯机制实例发布最终输出，则消耗的隐私预算总量将为

$$\sqrt{(r^* + 1)\mu_1^2 + (r^* + 1)\mu_2^2 + \mathcal{C}^2\mu_3^2},$$

我们需要通过求解下式来找到能够让我们“重新收回”未使用隐私预算的参数 C:

$$\sqrt{(r^* + 1)\mu_1^2 + (r^* + 1)\mu_2^2 + \mathcal{C}^2\mu_3^2} = \sqrt{R\mu_1^2 + R\mu_2^2 + \mu_3^2}.$$

该方程的解很容易得到:

$$\mathcal{C} = \frac{\sqrt{(R - r^* - 1)\mu_1^2 + (R - r^* - 1)\mu_2^2 + \mu_3^2}}{\mu_3} = \sqrt{(R - r^* - 1)\frac{\mu_1^2 + \mu_2^2}{\mu_3^2} + 1},$$

这样我们就可以在第 7 行的 if 语句中将 μ_3 按此因子放大，同时仍能实现所需的隐私保证。特别要注意的是，如果 $r^* = R - 1$，那么 $\mathsf{C} = 1$，并且不会

重新收回隐私预算；但对于任何更小的 r^* 值，只要 μ_1 或 $\mu_2 > 0$，就会导致隐私预算被重新收回。

□

B.1. **定理 3.1 的证明**。

证明。 根据命题 B.1，算法 3 的输出，即在$r^* = R$时的 $(\hat{m}_0, \ldots, \hat{m}_{r^*-1}, \hat{m}_{\mathsf{DP}}, r^*)$，或在 $r^* < R$ 时的 $(\hat{m}_0, \ldots, \hat{m}_{r^*}, \hat{m}_{\mathsf{DP}}, r^*)$，满足 $\sqrt{R\mu_1^2 + R\mu_2^2 + \mu_3^2}$-高斯差分隐私。此外，算法 1 是算法 3 在 $\mu_1 = \mu_2 = \frac{\mu}{2\sqrt{R}}$和 $\mu_3 = \frac{\mu}{\sqrt{2}}$下的一个实例。因此，$(\hat{m}_0, \ldots, \hat{m}_{r^*-1}, \hat{m}_{\mathsf{DP}}, r^*)$满足 μ-高斯差分隐私。此外， B^* 是对 r^*的确定性后处理。因此，算法 1 的输出满足 μ-高斯差分隐私。接下来，我们证明定理中的效用部分。在适当情况下，我们有时会使用标签 $\mu_1 = \mu_2 = \frac{\mu}{2\sqrt{R}}$ 和 $\mu_3 = \frac{\mu}{\sqrt{2}}$。

在继续证明之前，我们先定义几个事件，以便在下文中建立我们的结果。令 Z_r 为算法 3 第 6 行中的独立同分布标准高斯变量，并将事件 E_1 定义为

$$\mathcal{E}_1 = \Big\{ \max_{r \in [R]} |Z_r| \le \sqrt{2 \log \frac{4R}{\xi}} \Big\}, \tag{15}$$

根据引理 A.2，该事件以至少 $1 - \xi/2$ 的概率成立。令 N_r为算法 3 第 20 行中使用的独立同分布 $\mathsf{N}(0, I)d$ 维标准高斯变量，并将事件 E_2 定义为

$$\mathcal{E}_2 = \Big\{ \max_{r \in 0, \ldots, R-1} \|N_r\| \le \sqrt{d} + \sqrt{2 \log \frac{3(R+1)}{\xi}} \Big\}, \tag{16}$$

根据引理 A.3，该事件以至少 $1-\xi/3$ 的概率成立。最后，令 N_* 为算法 3 第 10 行或第 15 行中使用的终止高斯噪声，并将事件 E_3 定义为

$$\mathcal{E}_3 = \|N_*\| \le \sqrt{d} + \sqrt{2 \log \frac{6}{\xi}}, \tag{17}$$

该事件以至少 $1 - \xi/6$ 的概率成立。因此，事件 $\mathrm{E} = \mathrm{E}_1 \cap \mathrm{E}_2 \cap \mathrm{E}_3$ 以至少 $1 - \xi$ 的概率成立。我们在事件 E 条件下分析算法的性能。为此，我们将用到以下辅助引理，并在本定理的证明之后给出其证明。

引理 B.2. 假设定理 *3.1* 的假设成立。在由(15)-(17)构造的事件$\mathcal{E}$上，算法 *1* 的内部量满足以下性质：

性质 ***1***： $n_{LB} \ge n/2$。性质 ***2***： $\frac{B}{2^{r^*}} < 2K_2$。—

性质 ***3***： $\hat{m}_{\mathsf{DP}} = \frac{1}{|S|} \sum_{i \in S} D_i + B^* Z$ 其中 $|S| \ge n_{LB}$，对于 Z，一个在$\mathcal{E}$上有界的标准正态随机向量，且 $B^* = O(\frac{K_2}{\mu n})$。—

性质 ***4.*** 对于某个常数 $C > 0$ 和任意向量 $v \in \mathbb{R}^d$，以及最终估计 $\hat{m}_{\mathsf{DP}}$ 中包含的用户终止步索引集 S，

$$\Big\| \frac{1}{|S|} \sum_{i \in S} D_i - v \Big\| \le \|\bar{D} - v\| + \frac{CK_2}{n\mu} \sqrt{R \log \frac{R}{\xi}},$$

因此，

$$\|\hat{m}_{\mathsf{DP}} - v\| \le \|\bar{D} - v\| + \frac{CK_2}{n\mu}\sqrt{R\log\frac{R}{\xi}} + \frac{CK_2}{n\mu}\sqrt{d + \log\frac{1}{\xi}}.$$

现在可以直接证明定理 3.1 的效用结论。根据引理 B.2 的性质 4，取 $v = m$ 并根据假设使用 $\|D - m\| \le K_1$，即可立即得出

(18)

$$\|\hat{m}_{\mathsf{DP}} - m\| \le K_1 + \frac{CK_2}{n\mu}\sqrt{R\log\frac{R}{\xi}} + \frac{CK_2}{n\mu}\sqrt{d + \log\frac{1}{\xi}} = O\Big(K_1 + \frac{K_2}{n\mu}\sqrt{R\log\frac{R}{\xi}} + \frac{K_2}{n\mu}\sqrt{d + \log\frac{1}{\xi}}\Big).$$

□

引理 *B.2* 的证明。

证明。 我们证明，在事件 E 上，性质 1–3 确定性地成立，首先从性质 1 开始。

性质 1。 注意到 $2\tau - n = n - \sqrt[4]{\mu}\sqrt{2R\log\frac{4R}{\xi}}$，因此若 $n \ge \sqrt[8]{\mu}\sqrt{2R\log\frac{4R}{\xi}}$（这是我们的假设），则 $2\tau - n \ge n/2 \ge 1$，从而 $n_{LB} = 2\tau - n \ge n/2$。

性质 2。 根据 E_1,，我们知道若 $\sum_{i=1}^n \omega_i = n$，则

$$\omega \ge \sum_{i=1}^n \omega_i + \frac{2\sqrt{R}}{\mu}\min_{r\in[R]} Z_r \ge n - \frac{2}{\mu}\sqrt{2R\log\frac{4R}{\xi}} = \tau.$$

换言之，只要 $\sum_{i=1}^n \omega_i = n$，算法就不会终止。

接下来，我们将证明，在 E 下，只要对某个绝对常数 $C > 0$ 有 $n \ge \frac{C}{\mu}\sqrt{R(\log\frac{R}{\xi} + d)}$（如假设所述），算法就不会在第 r 步终止（如果 $B/2^r \ge 4K_2$）。因此，如果算法在第 $r < R$ 步提前终止，我们可以得知 $B/2^r < 4K_2$。在这种情况下，$r^* = r - 1$，因此 $B/2^{r^*} = B/2^{r-1} < 2K_2$。

现在我们通过归纳法证明这一点。具体而言，我们将证明，如果在第 $r-1 \ne R$ 步有 $\sum_{i=1}^n \omega_i = n$，那么在第 r 步有 $\sum_{i=1}^n \omega_i = n$，这意味着只要对某个绝对常数 $C > 0$ 有 $n \ge \frac{C}{\mu}\sqrt{R(\log\frac{R}{\xi} + d)}$ 且 $B/2^{r+1} \ge 2K_2$，算法就不会终止。

I在第零次迭代中，根据我们的假设 $\max_i \|D_i\| \le B$，我们有

$$\sum_{i=1}^n \omega_i = \sum_{i=1}^n \mathbb{1}\{\|D_i - \hat{m}_{-1}\| \le B\} = \sum_{i=1}^n \mathbb{1}\{\|D_i\| \le B\} = n.$$

现在假设在步骤 $r-1 \ne R$ 中 $\sum \omega_i = n$。在这种情况下，算法进入第 15 行，我们注意到 $|S| = \sum \omega_i = n$，这意味着 $n_{\tilde{S}} = |S|$，由此可得 $\frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r-2}) = \frac{1}{n}\sum_{i=1}^n (D_i - \hat{m}_{r-2})$。利用第 16 行中的 $\bar{D} := \frac{1}{n}\sum_{i=1}^n D_i$，我们得到

$$\hat{m}_{r-1} = \hat{m}_{r-2} + \frac{1}{n}\sum_{i=1}^n (D_i - \hat{m}_{r-2}) + \frac{4B\sqrt{R}}{2^{r-1}\mu n_{LB}}N_{r-1} = \bar{D} + \frac{4B\sqrt{R}}{2^{r-1}\mu n_{LB}}N_{r-1}.$$

现在我们希望找出使 $\sum_{i=1}^{n}\omega_i = n$ 在步骤 r 中也成立的条件。首先注意，在事件 E_2, 上

(19)
$$\|D_i - \hat{m}_{r-1}\| = \left\|D_i - \bar{D} - \frac{4B\sqrt{R}}{2^{r-1}\mu n_{LB}}N_{r-1}\right\| \le \|D_i - \bar{D}\| + \frac{4B\sqrt{R}}{2^{r-1}\mu n_{LB}}\|N_{r-1}\|$$
$$\le \|D_i - \bar{D}\| + \frac{4B\sqrt{R}}{2^{r-1}\mu n_{LB}}\left(\sqrt{d} + \sqrt{2\log\frac{3(R+1)}{\xi}}\right).$$

¯接下来，如果 **(A)** $\max_i \|D_i - D\| \le B/2^{r+1}$ 且 **(B)**

$$\frac{4B\sqrt{R}}{2^{r-1}\mu n_{LB}}\left(\sqrt{d} + \sqrt{2\log\frac{3(R+1)}{\xi}}\right) \le \frac{B}{2^{r+1}},$$

那么（19）对所有 i均推出 $\|D_i - \hat{}m_{r-1}\| \le B/2^r$ 。由此可见，在第 r 步中，我们再次有 $\sum\omega_i = n$。现在我们来寻找 **(A)** 和 **(B)** 成立的条件。

首先，对于条件**(A)**，根据假设，

$$\|D_i - \bar{D}\| = \|D_i - m - \bar{D} + m\| \le \|D_i - m\| + \|\bar{D} - m\| \le K_1 + K_2 \le 2K_2;$$

¯因此，$\max_i \|D_i - D\| \le 2K_2$。我们注意到，最后一步使用了 $K_1 \le K_2$，由于 $\max_i \|D_i - m\| \le K_2$ 蕴含 $\|D - m\| \le K_2$，这显然¯ ¯成立。因此，如果$2K_2 \le B/2^{r+1}$，则 max $_i \|D_i - D\| \le B/2^{r+1}$。

B由性质 2 可知 $n_{LB} = 2\tau - n$。因此，条件 **（B）** 得以满足 if

(20)
$$\frac{16\sqrt{R}}{\mu}\left(\sqrt{d} + \sqrt{2\log\frac{3(R+1)}{\xi}}\right) \le n_{LB} = 2\tau - n = n - \frac{4}{\mu}\sqrt{2R\log\frac{4R}{\xi}},$$

对于某个绝对常数 $C > 0$, 当 $n \ge \frac{C}{\mu}\sqrt{R(\log\frac{R}{\xi} + d)}$ 时该结论成立；根据假设，这一条件得以满足。

因此，我们已经证明：如果 n 足够大，除非半径足够小，否则算法不会提前终止：

(21)
$$B/2^{r^*} < 2K_2;$$

从而界定了诱导的隐私误差。此外，我们注意到，根据假设 $R \ge C\log(B/K_2)$，对于某个绝对常数 $C > 0$；这确保了在算法未提前终止、因此 $r^* = R$ 的情况下，我们有 $B/2^{r^*} = B/2^R \le K_2$.

性质 3。注意，如果我们在 r^* 处终止，其中 $r^* = R$ 或 $r^* < R$，那么当半径等于 $B/2^{r^*}$ 且中心为 $\hat{}m_{r^*-1}$ 时，算法通过了 ω 检查；我们将此标记为 ω^*，并注意到，由于 ω^* 通过了直线检查，我们必然有 $\omega^* \ge \tau$。

此外，为方便起见，以下我们令

$$\tilde{\mathcal{C}} := \mathcal{C}\mathbb{I}\{r^* < R\} + \mathbb{I}\{r^* = R\} = \mathbb{I}\{r^* < R\}\sqrt{2 - \frac{r^*+1}{R}} + \mathbb{I}\{r^* = R\} \ge 1,$$

以容纳两类终止事件。

现在我们考虑终止步骤中的 S，并注意到在 $r^* = R$ 或 $r^* = r-1 < R$ 任一情况下，都有 $|S| + \frac{2\sqrt{R}}{\mu} Z_{r^*} = \omega^*$。因此，在给定 E_1 的条件下，我们有

$$|S| = \omega^* - \frac{2\sqrt{R}}{\mu} Z_{r^*} \geq \tau - \frac{2}{\mu}\sqrt{2R\log\frac{4R}{\xi}} = n - \frac{4}{\mu}\sqrt{2R\log\frac{4R}{\xi}} = 2\tau - n = n_{LB}. \tag{22}$$

因此，直接可得，在事件 E_1 上，我们有 $n_{\tilde{S}} = \max\{|S|, n_{LB}\} = |S|$。由于根据 (22)，我们有 $\hat{m}_{\mathsf{DP}} = \hat{m}_{r^*-1} + \frac{1}{n_{\tilde{S}}}\sum_{i\in S}(D_i - \hat{m}_{r^*-1}) + \frac{2B}{2^{r^*}\tilde{\mathcal{C}}\mu_3 n_{LB}} N_*$，其中 $n_{\tilde{S}} = |S|$，因此可得 $\hat{m}_{\mathsf{DP}} = \frac{1}{|S|}\sum_{i\in S} D_i + \frac{2B}{2^{r^*}\tilde{\mathcal{C}}\mu_3 n_{LB}} N_*$。最后，对于 $B^* = \frac{2B}{2^{r^*}\tilde{\mathcal{C}}\mu_3 n_{LB}}$，注意到 $B/2^{r^*} \leq 2K_2$ 根据性质 2，$n_{LB} \geq n/2$ 根据性质 1，并且按构造 $\mathsf{C} \geq 1$ 和 $\mu_3 = \frac{\mu}{\sqrt{2}}$，因此 $B^* = O(\frac{K_2}{n\mu})$。

性质 4. 将最终输出记为 $\hat{m}_{\mathsf{DP}}$，并在事件 $\mathsf{E}_1 \cap \mathsf{E}_2$ 上分析我们的算法，可得

$$\|\hat{m}_{\mathsf{DP}} - v\| = \Big\|\frac{1}{|S|}\sum_{i\in S} D_i + \frac{2\sqrt{2}B}{2^{r^*}\tilde{\mathcal{C}}\mu n_{LB}} N_* - v\Big\| \leq \Big\|\frac{1}{|S|}\sum_{i\in S} D_i - v\Big\| + \frac{2\sqrt{2}B}{2^{r^*}\tilde{\mathcal{C}}\mu n_{LB}}\|N_*\|. \tag{23}$$

接下来，我们分别研究 (23) 右侧的两项。

我们首先证明 (23) 右侧的第二项为 $O(\frac{K_2}{n\mu}\sqrt{d + \log\frac{1}{\xi}})$。回顾性质 3，$B^* = O(\frac{K_2}{n\mu})$。因此，在事件 E 上，我们有

$$\frac{2\sqrt{2}B}{2^{r^*}\mu n_{LB}}\|N_*\| \leq \frac{CK_2}{n\mu}\Big(\sqrt{d + \log\frac{1}{\xi}}\Big). \tag{24}$$

现在考虑 (23) 右侧的第一项。记 $\bar{D}_S := \frac{1}{|S|}\sum_{i\in S} D_i$ 和 $S^{\mathsf{c}} := [n]\setminus S$。现在，由于 $nD = |S|D_S + |S^{\mathsf{c}}|D_{S^{\mathsf{c}}}$，可得

$$\bar{D}_S = \frac{n}{|S|}\bar{D} - \frac{|S^{\mathsf{c}}|}{|S|}\bar{D}_{S^{\mathsf{c}}} = \bar{D} - \frac{|S^{\mathsf{c}}|}{|S|}\left(\bar{D}_{S^{\mathsf{c}}} - \bar{D}\right); \tag{25}$$

因此，我们有

$$\Big\|\frac{1}{|S|}\sum_{i\in S} D_i - v\Big\| = \|\bar{D}_S - v\| = \Big\|\left(\bar{D} - v\right) - \frac{|S^{\mathsf{c}}|}{|S|}\left(\bar{D}_{S^{\mathsf{c}}} - \bar{D}\right)\Big\| \leq \left\|\bar{D} - v\right\| + \frac{|S^{\mathsf{c}}|}{|S|}\left\|\bar{D}_{S^{\mathsf{c}}} - \bar{D}\right\|.$$

第一项符合要求；对于第二项，我们利用

$$\left\|\bar{D}_{S^{\mathsf{c}}} - \bar{D}\right\| = \Big\|\frac{1}{|S^{\mathsf{c}}|}\sum_{i\in S^{\mathsf{c}}} D_i - \bar{D}\Big\| \leq \frac{1}{|S^{\mathsf{c}}|}\sum_{i\in S^{\mathsf{c}}}\left\|D_i - \bar{D}\right\| \leq \frac{1}{|S^{\mathsf{c}}|}\sum_{i\in S^{\mathsf{c}}}(\|D_i - m\| + \left\|\bar{D} - m\right\|) \leq 2K_2. \tag{26}$$

最后，为了证明所需的界，我们利用

$$\frac{|S^{\mathsf{c}}|}{|S|} = \frac{n - |S|}{|S|} \leq \frac{n - n_{LB}}{n_{LB}} \leq \frac{\frac{8}{\mu}\sqrt{2R\log\frac{4R}{\xi}}}{n} = O\Big(\frac{1}{n\mu}\sqrt{R\log\frac{R}{\xi}}\Big), \tag{27}$$

其中，第一个不等式由 (22) 得出，第二个不等式则由性质 1 得出。

由 (25)-(27) 可得

$$\Big\|\frac{1}{|S|}\sum_{i\in S} D_i - v\Big\| \le \|\bar{D} - v\| + \frac{CK_2}{n\mu}\sqrt{R\log\frac{R}{\xi}},$$

对于某个常数 $C > 0$，因此，结合 (24)，

$$\|\hat{m}_{\mathsf{DP}} - v\| \le \|\bar{D} - v\| + \frac{CK_2}{n\mu}\sqrt{R\log\frac{R}{\xi}} + \frac{CK_2}{n\mu}\sqrt{d + \log\frac{1}{\xi}}$$

□

B.2. **引理 3.2 的证明**。

证明。 关于 $\|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,\ t} - m\|$ 的陈述可直接由引理 A.1 与引理 A.4 结合得到。具体而言，对于 $u \in \mathbb{S}^{d-1}$， $u^{\mathsf{T}}(w_{i,\ t} - m)$ 是 σ^2-次高斯。因此，根据引理 A.1，

$$\mathbb{P}\Big(\Big|u^{\mathsf{T}}\Big[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,t} - m\Big]\Big| > q\Big) \le 2\exp\Big(-\frac{nTq^2}{2\sigma^2}\Big), \tag{28}$$

因此，将 (28) 的右侧设为 ξ 并求解 q，得到

$$\mathbb{P}\Big(\Big|u^{\mathsf{T}}\Big[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,t} - m\Big]\Big| > \sigma\sqrt{\frac{2}{nT}\log\frac{2}{\xi}}\Big) \le \xi.$$

接着，根据引理 A.4，

$$\mathbb{P}\Big(\Big\|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{i,t} - m\Big\| > \frac{4\sigma}{3}\sqrt{\frac{2}{nT}\Big(d\log 9 + \log\frac{2}{\xi}\Big)}\Big) \le \xi,$$

因此，该情形下的结论随之成立。对于 $\max_{i\in[n]}\|\frac{1}{T}\sum_{t=1}^{T} w_{i,t} - m\|$，注意到按照前述步骤，对于任意 $i \in [n]$，我们可以得到，

$$\mathbb{P}\Big(\Big\|\frac{1}{T}\sum_{t=1}^{T} w_{i,t} - m\Big\| > \frac{4\sigma}{3}\sqrt{\frac{2}{T}\Big(d\log 9 + \log\frac{2n}{\xi}\Big)}\Big) \le \xi/n,$$

因此，对 $[n]$ 使用并集界即可得到该情形下的结论。 □

B.3. **推论 3.3 的证明**。

证明。 令 E 表示引理 3.2 的界以概率参数$\xi' = \xi/2$ 成立的事件。在 E 上，根据定理 3.1 的缩放条件，定理 3.1 给出其成立概率至少为 $1 - \xi/2$，

$$\|\hat{m}_{\mathsf{DP}} - m\| = O\Big(\sigma\sqrt{\frac{1}{nT}\Big(d + \log\frac{1}{\xi}\Big)} + \sigma\sqrt{\frac{1}{n^2\mu^2 T}\Big(d + \log\frac{n}{\xi}\Big)\Big(d + \log\frac{1}{\xi} + R\log\frac{R}{\xi}\Big)}\Big),$$

因此，通过对这两个事件应用并集界，它们同时成立的概率至少为$1-\xi/2$；再注意到$\log\frac{1}{\xi}\le R\log\frac{R}{\xi}$，我们得到其成立概率至少为$1-\xi$，

$$\|\hat{m}_{\mathsf{DP}}-m\|=O\Big(\sigma\sqrt{\frac{1}{nT}\Big(d+\log\frac{1}{\xi}\Big)}+\sigma\sqrt{\frac{1}{n^2\mu^2T}\Big(d+\log\frac{n}{\xi}\Big)\Big(d+\log\frac{1}{\xi}+R\log\frac{R}{\xi}\Big)}\Big).$$

□

Appendix C. Proofs of Section 4

C.1. 引理 4.3 的证明。

证明。 我们首先推导一个事件，在该事件下$\mathsf{Rank}(X_i)=d$，从而$X_i^{\mathsf{T}}X_i$可逆且$\hat{}\ \beta_i=(X_i^{\mathsf{T}}X_i)^{-1}X_i^{\mathsf{T}}Y_i$。回顾$\Gamma_i:=\mathbb{E}[\frac{1}{T}X_i^{\mathsf{T}}X_i]$，根据假设 4.1，我们有$\lambda_{\min}(\Gamma_i)\ge c_\Gamma>0$。定义事件$\mathsf{E}_i:=\{\|\frac{1}{T}\Gamma_i^{-\frac{1}{2}}X_i^{\mathsf{T}}X_i\Gamma_i^{-\frac{1}{2}}-I\|\le\frac{3}{4}\}$。注意，在$\mathsf{E}_i$上，对于任意$v\neq\mathbf{0}\in\mathbb{R}^d$，

$$v^{\mathsf{T}}\frac{1}{T}X_i^{\mathsf{T}}X_iv=v^{\mathsf{T}}\Gamma_i^{\frac{1}{2}}\Big(\frac{1}{T}\Gamma_i^{-\frac{1}{2}}X_i^{\mathsf{T}}X_i\Gamma_i^{-\frac{1}{2}}\Big)\Gamma_i^{\frac{1}{2}}v=(\Gamma_i^{\frac{1}{2}}v)^{\mathsf{T}}\Big(\frac{1}{T}\Gamma_i^{-\frac{1}{2}}X_i^{\mathsf{T}}X_i\Gamma_i^{-\frac{1}{2}}\Big)\Gamma_i^{\frac{1}{2}}v>0,$$

因为在E_i上，$\lambda_{\min}(\frac{1}{T}\Gamma_i^{-\frac{1}{2}}X_i^{\mathsf{T}}X_i\Gamma_i^{-\frac{1}{2}})\ge\frac{1}{4}$，且$\Gamma_i^{\frac{1}{2}}$可逆，因此$\Gamma_i^{\frac{1}{2}}v\neq\mathbf{0}$。所以$\frac{1}{T}X_i^{\mathsf{T}}X_i$在$\mathsf{E}_i$上可逆。现在我们推导$\mathsf{E}_i$以高概率成立的条件。

令$z_{i,t}:=\Gamma_i^{-1/2}x_{i,t}$，根据假设 4.1，它们是代理方差为$\sigma_x^2$的次高斯变量。现在，对于任意的$u\in\mathbb{S}^{d-1}$，令$(W(u)_{i,t})_{t\ge1}:=(u^{\mathsf{T}}z_{i,t}z_{i,t}^{\mathsf{T}}u-\|u\|^2)_{t\ge1}$。根据引理 A.8，我们有$\mathbb{P}(|W(u)_{i,t}|>q)\le2e^{-C\sigma_x^2}$对所有$i\in[n]$均成立，并且利用$\mathbb{E}[u^{\mathsf{T}}z_{i,t}z_{i,t}^{\mathsf{T}}u]=u^{\mathsf{T}}\Gamma_i^{-1/2}\mathbb{E}[x_{i,t}x_{i,t}^{\mathsf{T}}]\Gamma_i^{-1/2}u=u^{\mathsf{T}}u$可知$t\in[T]$。此外，根据引理 A.10，由于$z_{i,t}$是指数强混合的，因此序列$(W(u)_{i,t})_{t\in\mathbb{Z}}$对所有$i\in[n]$也是指数强混合的。因此，应用引理 A.14 并取$K:=C\sigma_x^2$，我们以至少$1-\xi$的概率得到，

$$\Big|\frac{1}{T}\sum_{t=1}^{T}W(u)_{i,t}\Big|=O\Big(\sigma_x^2\sqrt{\frac{1}{T}\log\frac{1}{\xi}}+\frac{C_{\sigma_x}}{T}\big(\log\frac{T}{\xi}\big)^2\Big).\tag{29}$$

利用ϵ-网论证（引理 A.4）可得，以至少$1-\xi$的概率，

$$\Big\|\frac{1}{T}\sum_{t=1}^{T}z_{i,t}z_{i,t}^{\mathsf{T}}-I\Big\|=O\Big(\sigma_x^2\sqrt{\frac{1}{T}\Big(d+\log\frac{1}{\xi}\Big)}+\frac{C_{\sigma_x}}{T}\Big(d+\log\frac{T}{\xi}\Big)^2\Big).$$

特别地，对于任意$i\in[n]$，只要$T\ge C(d+\log\frac{nT}{\xi})^2$，其中$C>0$为某个常数，

$$\mathbb{P}(\mathcal{E}_i^{\mathsf{c}})=\mathbb{P}\left(\Big\|\frac{1}{T}\sum_{t=1}^{T}z_{i,t}z_{i,t}^{\mathsf{T}}-I\Big\|>\frac{3}{4}\right)=\mathbb{P}\left(\Big\|\frac{1}{T}\Gamma_i^{-\frac{1}{2}}X_i^{\mathsf{T}}X_i\Gamma_i^{-\frac{1}{2}}-I\Big\|>\frac{3}{4}\right)\le\frac{\xi}{2n}.\tag{30}$$

因此，只要$T\ge C(d+\log\frac{nT}{\xi})^2$，，$\mathsf{E}_i$就以至少$1-\frac{\xi}{2n}$的概率成立；我们将在此事件上分析误差界。

$\hat{}$我们从$\max_{i\in[n]}\|\beta_i-\beta\|$的集中性开始。首先证明，对于任意$i\in[n]$和$u\in\mathbb{S}^{d-1}$，以及$\mathbb{R}^d$中的单位球，$\hat{}u^{\mathsf{T}}(\beta_i-\beta)$具有集中性，然后通过并集界逐步

逼近我们的目标。为此，对于任意 $i \in [n]$，注意到在 E_i 上，我们有 $\hat{}\beta_i = (X_i^\mathsf{T} X_i)^{-1} X_i^\mathsf{T} Y_i$，

(31)
$$\begin{aligned}
\mathbb{P}(|u^\mathsf{T}(\hat{\beta}_i - \beta)| > q) &\le \mathbb{E}_{X_i}\left[\mathbb{1}_{\mathcal{E}_i}\mathbb{P}\left(|u^\mathsf{T}(\hat{\beta}_i - \beta)| > q \mid X_i\right)\right] + \mathbb{P}(\mathcal{E}_i^\mathsf{c}) \\
&= \mathbb{E}_{X_i}\left[\mathbb{1}_{\mathcal{E}_i}\mathbb{P}\left(|u^\mathsf{T}(X_i^\mathsf{T} X_i)^{-1} X_i^\mathsf{T}\varepsilon_i| > q \mid X_i\right)\right] + \mathbb{P}(\mathcal{E}_i^\mathsf{c}) \\
&= \mathbb{E}_{X_i}\Big[\mathbb{1}_{\mathcal{E}_i}\mathbb{P}\Big(\Big|\frac{u^\mathsf{T}(X_i^\mathsf{T} X_i)^{-1} X_i^\mathsf{T}}{\|u^\mathsf{T}(X_i^\mathsf{T} X_i)^{-1} X_i^\mathsf{T}\|}\varepsilon_i\Big| > \frac{q}{\|u^\mathsf{T}(X_i^\mathsf{T} X_i)^{-1} X_i^\mathsf{T}\|}\,\Big|\, X_i\Big)\Big] + \mathbb{P}(\mathcal{E}_i^\mathsf{c}) \\
&\le \mathbb{E}_{X_i}\Big[\mathbb{1}_{\mathcal{E}_i}\mathbb{P}\Big(|\tilde{u}^\mathsf{T}\varepsilon_i| > \frac{q}{C_i}\,\Big|\, X_i\Big)\Big] + \mathbb{P}(\mathcal{E}_i^\mathsf{c}),
\end{aligned}$$

其中 $C_i := \|(X_i^\mathsf{T} X_i)^{-1} X_i^\mathsf{T}\| \ge \|u^\mathsf{T}(X_i^\mathsf{T} X_i)^{-1} X_i^\mathsf{T}\|$ 且 $\tilde{u} \in \mathbb{S}^{T-1}$。现在，考察 (31)，我们注意到，根据定义 2.8 和假设 4.2，$\tilde{u}^\mathsf{T}\varepsilon_i$ 是一个 σ_ε^2-次高斯随机变量，因此可以应用引理 A.1 的尾界（取 $n = 1$），得到

$$\mathbb{P}\left(\left|u^\mathsf{T}(\hat{\beta}_i - \beta)\right| > q\right) \le \mathbb{E}_{X_i}\left[\mathbb{1}_{\mathcal{E}_i} 2\exp\left\{\frac{-q^2}{2\sigma_\varepsilon^2 C_i^2}\right\}\right] + \mathbb{P}(\mathcal{E}_i^\mathsf{c}). \tag{32}$$

现在的关键是在 E_i 下推导 C_i 的一个确定性界。我们有，

$$\begin{aligned}
C_i = \|(X_i^\mathsf{T} X_i)^{-1} X_i^\mathsf{T}\| &= \|\Gamma_i^{-\frac{1}{2}}\Gamma_i^{\frac{1}{2}}(X_i^\mathsf{T} X_i)^{-\frac{1}{2}}(X_i^\mathsf{T} X_i)^{-\frac{1}{2}} X_i^\mathsf{T}\| \\
&\le \frac{1}{\sqrt{T}}\|\Gamma_i^{-\frac{1}{2}}\| \cdot \left\|\Gamma_i^{\frac{1}{2}}\Big(\frac{1}{T}X_i^\mathsf{T} X_i\Big)^{-\frac{1}{2}}\right\| \cdot \|(X_i^\mathsf{T} X_i)^{-\frac{1}{2}} X_i^\mathsf{T}\|.
\end{aligned} \tag{33}$$

首先，根据引理 A.6，在 E_i 上，我们有，

$$\left\|\Gamma_i^{\frac{1}{2}}\Big(\frac{1}{T}X_i^\mathsf{T} X_i\Big)^{-\frac{1}{2}}\right\| \le \frac{1}{\sqrt{1 - 3/4}} = 2, \tag{34}$$

(33) 中的第三项在 E_i 上等于 1，因为 $(X_i^\mathsf{T} X_i)^{-\frac{1}{2}} X_i^\mathsf{T} X_i (X_i^\mathsf{T} X_i)^{-\frac{1}{2}} = I$. 因此，利用这一点以及 (34) 代入 (33)，可得在 ξ_i 上有

$$C_i^2 \le \frac{1}{T}\left\|\Gamma_i^{\frac{1}{2}}\Big(\frac{1}{T}X_i^\mathsf{T} X_i\Big)^{-\frac{1}{2}}\right\|^2 \|\Gamma_i^{-\frac{1}{2}}\|^2 \le \frac{4}{T}\|\Gamma_i^{-\frac{1}{2}}\|^2. \tag{35}$$

将此界代入（32），我们得到

(36)
$$\mathbb{P}\left(\left|u^\mathsf{T}(\hat{\beta}_i - \beta)\right| > q\right) \le 2\mathbb{E}_{X_i}\left[\mathbb{1}_{\mathcal{E}_i}\exp\left\{\frac{-q^2}{2\sigma_\varepsilon^2 C_i^2}\right\}\right] + \mathbb{P}(\mathcal{E}_i^\mathsf{c}) \le 2\exp\left\{\frac{-Tq^2}{8\sigma_\varepsilon^2 \max_j \|\Gamma_j^{-1/2}\|^2}\right\} + \mathbb{P}(\mathcal{E}_i^\mathsf{c}).$$

因此，根据（30）和（36），对于任意 $u \in \mathbb{S}^{d-1}$，当 $q \ge 2\sigma_\varepsilon \max_{1\le j\le n} \|\Gamma_j^{-\frac{1}{2}}\| \sqrt{\frac{2}{T}\log\frac{4n}{\xi}}$ 时，

$$\max_{i\in[n]} \mathbb{P}\left(\left|u^\mathsf{T}(\hat{\beta}_i - \beta)\right| > q\right) \le 2\exp\left\{\frac{-Tq^2}{8\sigma_\varepsilon^2 \max_i \|\Gamma_i^{-1/2}\|^2}\right\} + \frac{\xi}{2n} \le \frac{\xi}{n}. \tag{37}$$

W现在可以对 $[n]$ 应用并集界，并将引理 A.4 应用于（37），从而得到 n

$$\mathbb{P}\Big(\max_{i\in[n]}\big\|\hat{\beta}_i-\beta\big\|>\frac{4}{3}2\sigma_\varepsilon\max_{i\in[n]}\|\Gamma_i^{-\frac{1}{2}}\|\sqrt{\frac{2}{T}\log\frac{9^d4n}{\xi}}\Big)$$
$$\le n\max_{i\in[n]}\Big[\mathbb{P}\Big(\big\|\hat{\beta}_i-\beta\big\|>\frac{4}{3}2\sigma_\varepsilon\max_{i\in[n]}\|\Gamma_i^{-\frac{1}{2}}\|\sqrt{\frac{2}{T}\log\frac{9^d4n}{\xi}}\Big)\Big]\le n\frac{\xi}{n}=\xi,$$

由此得以完成证明：以至少 $1-\xi$ 的概率,

$$\max_{i\in[n]}\big\|\hat{\beta}_i-\beta\big\|=O\Big(\sigma_\varepsilon\max_{i\in[n]}\|\Gamma_i^{-\frac{1}{2}}\|\sqrt{\frac{1}{T}\big(d+\log\frac{n}{\xi}\big)}\Big).$$

¯我们继续证明定理的第二部分，该部分涉及 $\beta-\beta$ 的集中性。证明策略与上文类似。注意到，对于任意 $u\in\mathbb{S}^{d-1}$， $u^\intercal(\bar{\beta}-\beta)=u^\intercal(\frac{1}{n}\sum_{i=1}^n\hat{}\beta_i-\beta)=\frac{1}{n}\sum_{i=1}^n u^\intercal(\hat{}\beta_i-\beta)$。然后，如同式 (31) 和 (32) 中那样，现在使用引理 A.1，其中 n 个相互独立的 $\|u^\intercal(X_i^\intercal X_i)^{-1}X_i^\intercal\|^2\sigma_\varepsilon^2$-次高斯随机变量在给定 X 的条件下，并利用给定 X 时 ε_i 与 ε_j 相互独立这一点，并定义 $\mathrm{E}:=\cap_{i=1}^n\mathrm{E}_i$,

$$\mathbb{P}\left(\left|u^\intercal(\bar{\beta}-\beta)\right|>q\right)\le 2\mathbb{E}_X\Big[\mathbb{1}_{\mathcal{E}}\exp\Big\{\frac{-n^2q^2}{2\sigma_\varepsilon^2\sum_{i=1}^n C_i^2}\Big\}\Big]+\mathbb{P}(\mathcal{E}^{\mathrm{c}})\le 2\exp\Big\{\frac{-n^2Tq^2}{8\sigma_\varepsilon^2\sum_{i=1}^n\|\Gamma_i^{-1/2}\|^2}\Big\}+\frac{\xi}{2}\le\xi,$$

其中，最后一个不等式对.....成立

$$q=\Omega\Big(\sigma_\varepsilon\sqrt{\frac{1}{n}\sum_{i=1}^n\|\Gamma_i^{-1/2}\|^2}\sqrt{\frac{1}{nT}\log\frac{1}{\xi}}\Big).$$

因此，存在 $C'>0$，使得对所有 $u\in\mathbb{S}^{d-1}$,

$$\mathbb{P}\Big(\left|u^\intercal(\bar{\beta}-\beta)\right|>C\sigma_\varepsilon\sqrt{\frac{1}{n}\sum_{i=1}^n\|\Gamma_i^{-1/2}\|^2}\sqrt{\frac{1}{nT}\log\frac{1}{\xi}}\Big)\le\xi. \tag{38}$$

最后，再次将引理 A.4 应用于 (38)，我们得到，以至少 $1-\xi$ 的概率,

$$\|\bar{\beta}-\beta\|=O\Big(\sigma_\varepsilon\sqrt{\frac{1}{n}\sum_{i=1}^n\|\Gamma_i^{-1/2}\|^2}\sqrt{\frac{1}{nT}\Big(d+\log\frac{1}{\xi}\Big)}\Big).$$

□

C.2. **引理 4.5 的证明**。

证明。 对于任意 $i\in[n]$，令 $\Gamma_{i,T}:=\frac{1}{T}X_i^\intercal X_i$ 为所关注的对称半正定矩阵。根据算子范数的定义,

$$\mathbb{E}\left[\|\Gamma_{i,T}^{-1}\|^{k+\delta}\right]=\mathbb{E}\left[\lambda_{\min}(\Gamma_{i,T})^{-(k+\delta)}\right]=\int_0^\infty\mathbb{P}\big(\lambda_{\min}(\Gamma_{i,T})^{-(k+\delta)}>r\big)dr$$
$$=\int_0^\infty\mathbb{P}\big(\lambda_{\min}(\Gamma_{i,T})<r^{\frac{-1}{k+\delta}}\big)dr\le\int_0^\infty\mathbb{P}(\lambda_{\min}(\Gamma_{i,T})\le r^{\frac{-1}{k+\delta}})dr.$$

现在，通过变量替换，

$$\mathbb{E}\Big[\big\|\Gamma_{i,T}^{-1}\big\|^{k+\delta}\Big] \le \int_0^\infty \mathbb{P}\big(\lambda_{\min}(\Gamma_{i,T}) \le r^{\frac{-1}{k+\delta}}\big)dr = (k+\delta)\int_0^\infty r^{-(k+\delta)-1}\mathbb{P}\big(\lambda_{\min}(\Gamma_{i,T}) \le r\big)dr$$

$$\le C\Big(\int_0^{r_0} r^{-(k+\delta)-1}\mathbb{P}\big(\lambda_{\min}(\Gamma_{i,T}) \le r\big)dr + \frac{1}{r_0}\Big), \tag{39}$$

其中，最后一步对任意 $r_0 > 0$ 都成立。因此，为完成证明，只需找到一个$r_0 > 0$ 和 T_0，使得对于所有 $i \in [n]$ 和 $T \ge T_0$，都有

$$\int_0^{r_0} r^{-(k+\delta)-1}\mathbb{P}(\lambda_{\min}(\Gamma_{i,T}) \le r)dr < \infty. \tag{40}$$

为此，我们注意到，对于任意 u， $v \in \mathbb{S}^{d-1}$ 以及任意对称的、半正定的 $\Gamma_{i,\,T}$，

$$\left|u^\mathsf{T}\Gamma_{i,T}\,u - v^\mathsf{T}\Gamma_{i,T}\,v\right| = \left|(u-v)^\mathsf{T}\Gamma_{i,T}\,u + v^\mathsf{T}\Gamma_{i,T}\,(u-v)\right| \le 2\|\Gamma_{i,T}\|\cdot\|u-v\|.$$

现在令 $\mathbf{N}$ 为 ϵ的一个 $\mathbb{S}^{d-1}$-网，这意味着对于所有 $u \in \mathbb{S}^{d-1}$，都存在 $\tilde{u} \in \mathbf{N}$，使得$\|v-\tilde{u}\| \le \epsilon$。然后，令 $v_i :=$ 为$_{u\in\mathbb{S}^{d-1}}\, u^\mathsf{T}\Gamma_i$的极小值点$_{,\,T}\,u$，并令 $u_i \in \mathbf{N}$ 使得 $\|u_i - v_i\| \le \epsilon$。于是我们知道

$$\lambda_{\min}(\Gamma_{i,T}) = v_i^\mathsf{T}\Gamma_{i,T}\,v_i \ge u_i^\mathsf{T}\Gamma_{i,T}\,u_i - 2\|\Gamma_{i,T}\|\cdot\|u_i - v_i\| \ge u_i^\mathsf{T}\Gamma_{i,T}\,u_i - 2\epsilon\|\Gamma_{i,T}\|$$

$$\ge \min_{u\in\mathcal{N}} u^\mathsf{T}\Gamma_{i,T}\,u - 2\epsilon\|\Gamma_{i,T}\|.$$

因此，对于任意任意 $L > 0$,

$$\Big\{\lambda_{\min}(\Gamma_{i,T}) \le r\Big\} \subseteq \Big\{\min_{u\in\mathcal{N}} u^\mathsf{T}\Gamma_{i,T}\,u \le r + 2\epsilon\|\Gamma_{i,T}\|\Big\}$$

$$= \Big\{\min_{u\in\mathcal{N}} u^\mathsf{T}\Gamma_{i,T}\,u \le r + 2\epsilon\|\Gamma_{i,T}\|\Big\} \cap \Big(\big\{\|\Gamma_{i,T}\| \le L\big\} \cup \big\{\|\Gamma_{i,T}\| > L\big\}\Big)$$

$$\subseteq \Big\{\min_{u\in\mathcal{N}} u^\mathsf{T}\Gamma_{i,T}\,u \le r + 2\epsilon L\Big\} \cup \big\{\|\Gamma_{i,T}\| > L\big\}.$$

然后，对这些事件应用并集界，可得

$$\mathbb{P}\left(\lambda_{\min}(\Gamma_{i,T}) \le r\right) \le \mathbb{P}\Big(\min_{v\in\mathcal{N}} v^\mathsf{T}\Gamma_{i,T}\,v \le r + 2\epsilon L\Big) + \mathbb{P}\left(\|\Gamma_{i,T}\| > L\right). \tag{41}$$

回顾一下，我们要证明式（40）中的界。考虑式（41）右侧的第二项。根据式（12）和三角不等式，以至多为 ξ 的概率,

$$\|\Gamma_{i,T}\| > \|\Gamma_i\| + C_\Gamma\sigma_x^2\sqrt{\frac{1}{T}\Big(d + \log\frac{1}{\xi}\Big)} + \frac{C_{\sigma_x,\Gamma}}{T}\Big(\log\frac{T}{\xi}\Big)^2,$$

因此，若设 $L := C_\Gamma + C_\Gamma\sigma_x^2\sqrt{\frac{1}{T}(d+\log\, r^{-(k+\delta+\gamma)})} + \frac{C_{\sigma_x,\Gamma}}{T}(\log[Tr^{-(k+\delta+\gamma)}])^2$,，则 $\mathbb{P}(\|\Gamma_{i,\,T}\| > L) \le r^{k+\delta+\gamma}$。因此，对于这一 L 的选择，由 (41) 可得,

$$\mathbb{P}\left(\lambda_{\min}(\Gamma_{i,T}) \le r\right) \le \mathbb{P}\Big(\min_{v\in\mathcal{N}} v^\mathsf{T}\Gamma_{i,T}\,v \le r + 2\epsilon L\Big) + r^{k+\delta+\gamma}, \tag{42}$$

因此，为完成证明，我们给出 (42) 右侧第一项的上界。最后，设 $\epsilon = \frac{r}{2L}$，回顾一下 $0 < r \le r_0$。现在令 $r_0 \le e^{-1}$，接下来我们将对上文定义的 L 给出上界：首先注意到

$$C_\Gamma \sigma_x^2 \sqrt{\frac{1}{T}\Big(d - (k+\delta+\gamma)\log r\Big)} \le C_\Gamma \sigma_x^2 \sqrt{d + (k+\delta+\gamma)|\log r|} \le C\left(1 + |\log r|^2\right) \le C_1 |\log r|^2,$$

对于常数 C_1，$C > 0$，其不依赖于 r 和 T，而定义 $u := (k+\delta+\gamma)|\log r|$

$$\frac{C_{\sigma_x,\Gamma}}{T}\Big(\log[Tr^{-(k+\delta+\gamma)}]\Big)^2 \le \frac{C_{\sigma_x,\Gamma}}{T}(\log T + u)^2 \le C_{\sigma_x,\Gamma}(4 + 2u^2) \le c(1 + |\log r|^2) \le C_2 |\log r|^2,$$

对于常数 $c, C_2 > 0$，其不依赖于 r 和 T。由此可得，对于某个常数 $C_3 > 0$，$L^d \le C_3^d (\log r)^{2d}$，该常数不依赖于 r 和 T。利用 Vershynin（2018）的推论 4.2.13 以及关于 L 的上界，我们有 $|\mathbb{N}| \le (6L/r)^d \le 6^d C_3^d |\log r|^{2d} r^{-d} =: C' |\log r|^{2d} r^{-d}$。因此，对 $\mathbb{N}$ 使用并集界，

$$\mathbb{P}\Big(\min_{v\in\mathcal{N}} v^\top \Gamma_{i,T} v \le r + 2\epsilon L\Big) \le |\mathcal{N}| \sup_{v\in\mathcal{N}} \mathbb{P}(v^\top \Gamma_{i,T} v \le 2r) \le C' |\log r|^{2d} r^{-d} \sup_{v\in\mathcal{N}} \mathbb{P}(v^\top \Gamma_{i,T} v \le 2r), \tag{43}$$

因此，只需对任意 $v \in \mathbb{S}^{d-1}$ 控制 $\mathbb{P}(v^\top \Gamma_{i,T} v \le 2r)$。固定 $v \in \mathbb{S}^{d-1}$，并令该值为 $w_{i,t} := x_{i,t}^\top v$，利用关于 f_{W_i} 的假设上界，我们有

$$\mathbb{P}\Big(v^\top \Gamma_{i,T} v \le 2r\Big) = \mathbb{P}\Big(\sum_{t=1}^{T} (x_{i,t}^\top v)^2 < 2rT\Big) = \int_{\|w\| \le \sqrt{2rT}} f_{W_i}(w) dw \le K^T \int_{\|w\| \le \sqrt{2rT}} 1 dw.$$

根据上述内容以及不等式 $\Gamma(z) > \left(\frac{z}{e}\right)^z \sqrt{\frac{2\pi}{z}}$（例如 Olver 2010年），我们可知

$$\mathbb{P}\Big(v^\top \Gamma_{i,T} v \le 2r\Big) \le K^T \mathrm{Vol}(B_T(\sqrt{2rT})) = K^T \frac{\pi^{T/2}(\sqrt{2rT})^T}{\Gamma(T/2+1)} \le (4\pi K^2 e r)^{T/2}. \tag{44}$$

将 (43)-(44) 中的界代入 (42)，可得

$$\begin{aligned} \mathbb{P}\left(\lambda_{\min}(\Gamma_{i,T}) \le r\right) &\le C' |\log r|^{2d} r^{-d} (4\pi K^2 e r)^{T/2} + r^{k+\delta+\gamma} \\ &= C' (4\pi K^2 e)^d |\log r|^{2d} (4\pi K^2 e r)^{T/2-d} + r^{k+\delta+\gamma}, \end{aligned} \tag{45}$$

其中 C' 是一个不依赖于 r 和 T的常数。我们要利用 (45) 中的界证明 (40)。令 $r \le r_0 := \min\{\frac{1}{4\pi e K^2}, e^{-1}\}$，则有 $4\pi K^2 e r \le 1$，因此 $(4\pi K^2 e r)^{T/2-d}$关于~~T递~~减。现在令 $T_0 \ge 2d$ 为某个阈值，则对于 $T \ge T_0$，

$$(4\pi K^2 e r)^{T/2-d} \le (4\pi K^2 e r)^{T_0/2-d},$$

因此我们得到

$$\mathbb{P}(\lambda_{\min}(\Gamma_{i,t}) \le r) \le C' (4\pi K^2 e)^{T_0/2} |\log r|^{2d} r^{T_0/2-d} + r^{k+\delta+\gamma}, \tag{46}$$

另外注意到，由 (46) 可得

$$\mathbb{P}(\lambda_{\min}(\Gamma_{i,T}) = 0) = \lim_{r\to 0^+} \mathbb{P}(\lambda_{\min}(\Gamma_{i,T}) \le r) \le \lim_{r\to 0^+} \left(C' (4\pi K^2 e)^{T_0/2} |\log r|^{2d} r^{T_0/2-d} + r^{k+\delta+\gamma}\right) = 0,$$

从而对于所有 $T \ge T_0 > 2d$，逆矩阵几乎必然存在。

最后，将 (46) 代入 (39)，可得，对于任意 $T \geq T_0$,

$$\begin{aligned}\mathbb{E}[\|\Gamma_{i,T}^{-1}\|^{k+\delta}] &\leq C_1 + C\int_0^{r_0} r^{-(k+\delta)-1}\left(C'(4\pi K^2 e)^d |\log r|^{2d}(4\pi K^2 e r)^{T_0/2-d} + r^{k+\delta+\gamma}\right) dr \\ &\leq C_1 + C_2 + CC'(4\pi K^2 e)^{T_0/2}\int_0^{r_0} r^{-(k+\delta)-1+T_0/2-d}|\log r|^{2d} dr,\end{aligned}$$

且当且仅当 $-(k+\delta) - 1 + T_0/2 - d > -1$ 时，后一积分收敛；换言之，$T_0 > 2(d+k+\delta)$。因此，对于任意满足 $T_0 > 2(d+k+\delta)$ 的有限 T_0，我们便得到所需结论

$$\sup_{i\in[n], T\geq T_0} \mathbb{E}\Big[\Big\|\Big(\frac{1}{T}X_i^\intercal X_i\Big)^{-1}\Big\|^{k+\delta}\Big] < \infty.$$

□

C.3. **定理 4.6 的证明。**

Proof. 对于 (a) 部分，隐私保证可由定理 3.1 立即得出 .

对于 (b) 部分，误差界也可由定理 3.1 得出，只需按如下方式选取上界 B、 K_1 和 K_2： $B = \Omega(\|\beta\| + K_2)$，其中

(47)

$$K_1 = O\Big(\sigma_\varepsilon\sqrt{\frac{1}{n}\sum_{i=1}^n \|\Gamma_i^{-1/2}\|^2}\sqrt{\frac{1}{nT}\big(d + \log\frac{1}{\xi}\big)}\Big), \text{ and } K_2 = O\Big(\sigma_\varepsilon \max_{i\in[n]}\|\Gamma_i^{-1/2}\|\sqrt{\frac{1}{T}\big(d+\log\frac{n}{\xi}\big)}\Big).$$

事实上，引理 4.3 表明，这些 B、 K_1 和 K_2 的取值满足：以至少 $1-\xi$ 的概率，它们分别界定 $\max_i \|\beta_i\|$、 $\|\beta - \beta\|$ 和 $\max_i \|\beta_i - \beta\|$。因此，(b) 部分可由定理 3.1 的陈述以及引理 4.3 对我们提出的 B、 K_1 和 K_2 所保证的高概率保证得出。

定理的 (c) 部分涉及估量的渐近分布。我们通过证明在定理所述条件下，(i) $\sqrt{nT}(\hat{\beta}_{\mathsf{DP}} - \bar{\beta}) \xrightarrow{p} 0$，并且 (ii) $\sqrt{nT}(\bar{\beta} - \beta) \xrightarrow{d} N(0, \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})$，来证明这一点。最后，我们应用斯卢茨基定理得出结论。

我们先证明 (i)。首先，由假设 4.4 可知，对于给定的固定 (n, T) 对，其中$T \geq T_0$，有 $\mathbb{P}(\forall i \in [n] : 秩(X_i) = d) = 1$。现在考虑任意一个发散的双重序列 $(n_m, T_m) \to \infty$，其由 $m \in \mathbb{N}$ 索引。令 $m_0 = \min\{m : T_m \geq T_0\}$。那么 $\mathbb{P}(\forall m \geq m_0, i \in [n_m] : 秩(X_i) = d) = 1$，因为这是可数个几乎必然事件的交，因此仍然是几乎必然事件。因此， $\mathbb{P}(\forall m \geq m_0, i \in [n_m] : \beta_i = \beta_{i,\mathsf{OLS}}) = 1$，所以在下文中，我们可以分析估量在 $\hat{\beta}_i = \hat{\beta}_{i,\mathsf{OLS}} = (X_i^\intercal X_i)^{-1}X_i^\intercal Y_i$ 时的渐近性质。

现在，注意到根据引理 B.2 的性质 4 以及由 (47) 得到的 K_2 值，我们有，将性质 4 的陈述中的 v 替换为 β，

$$\begin{aligned}\sqrt{nT}\|\hat{\beta}_{\mathsf{DP}} - \bar{\beta}\| &\leq \sqrt{nT}\|\bar{\beta} - \bar{\beta}\| + \frac{CK_2}{n\mu}\sqrt{R\log\frac{R}{\xi}} + \frac{CK_2}{n\mu}\sqrt{d + \log\frac{1}{\xi}} \\ &= 0 + O\Big(\frac{\sigma_\varepsilon}{\sqrt{n}\mu}\max_{i\in[n]}\|\Gamma_i^{-\frac{1}{2}}\|\sqrt{(d+\log\frac{n}{\xi})(d + R\log\frac{R}{\xi})}\Big).\end{aligned}$$

由于 $\|\Gamma_i^{-1/2}\|$ 一致有界（假设 4.1），因此当 $\xi \to 0$ 时，我们有 $\sqrt{nT}(\widehat{\beta_{\mathsf{DP}}} - \bar{\beta}) \xrightarrow{p} 0$；而 $\frac{1}{n\mu^2}(R\log\frac{R}{\xi} + \log\frac{1}{\xi})(\log\frac{n}{\xi}) \to 0$. 由于 $R \geq 1$，在 (5) 中所述的增长条件下，我们有

$$\frac{1}{n\mu^2}\Big(R\log\frac{R}{\xi} + \log\frac{1}{\xi}\Big)\Big(\log\frac{n}{\xi}\Big) \leq \frac{2}{n}\Big(R\log\frac{R}{\xi}\Big)\Big(\log\frac{n}{\xi}\Big) = o(1).$$

因此，$\sqrt{nT}(\widehat{\beta}_{\mathsf{DP}} - \bar{\beta}) \xrightarrow{p} 0$。

现在我们继续证明 (ii)。我们有

$$\bar{\beta} = \frac{1}{n}\sum_{i=1}^{n}(X_i^\mathsf{T}X_i)^{-1}X_i^\mathsf{T}Y_i = \beta + \frac{1}{n}\sum_{i=1}^{n}(X_i^\mathsf{T}X_i)^{-1}X_i^\mathsf{T}\varepsilon_i.$$

定义 $\Gamma_{i,T} := \frac{1}{T}X_i^\mathsf{T}X_i$，以及 $S_{i,T} := \frac{1}{\sqrt{T}}X_i^\mathsf{T}\varepsilon_i$。那么，我们有

$$\sqrt{nT}(\bar{\beta} - \beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma_{i,T}^{-1}S_{i,T} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma_i^{-1}S_{i,T} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Gamma_{i,T}^{-1} - \Gamma_i^{-1})S_{i,T}, \tag{48}$$

我们首先证明 (48) 右侧的第二项为 $o_p(1)$。我们有

$$\mathbb{E}\Big[\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Gamma_{i,T}^{-1} - \Gamma_i^{-1})S_{i,T}\Big\|^2\Big] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[S_{i,T}^\mathsf{T}(\Gamma_{i,T}^{-1} - \Gamma_i^{-1})(\Gamma_{j,T}^{-1} - \Gamma_j^{-1})S_{j,T}] \tag{49}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|(\Gamma_{i,T}^{-1} - \Gamma_i^{-1})S_{i,T}\|^2] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\Gamma_i^{-1}(\Gamma_i - \Gamma_{i,T})\Gamma_{i,T}^{-1}S_{i,T}\|^2], \tag{50}$$

这里，我们利用 X_i 和 ε_i 在 i之间的截面独立性（假设 4.1 和 4.2）以及严格外生性，得到第二个等号；因为在这些假设下$\mathbb{E}[S_{i,T}^\mathsf{T}(\Gamma_{i,T}^{-1} - \Gamma_i^{-1})(\Gamma_{j,T}^{-1} - \Gamma_j^{-1})S_{j,T}] = 0$对于 $i \neq j$。然后，根据 $\|\Gamma_i^{-1}\|$（假设 4.1）的统一有界性和 Holder 不等式，我们可以从 (50) 推得，

$$\sup_{i\geq 1}\mathbb{E}[\|\Gamma_i^{-1}(\Gamma_i - \Gamma_{i,T})\Gamma_{i,T}^{-1}S_{i,T}\|^2] \leq \sup_{i\geq 1}\|\Gamma_i^{-1}\|^2 \cdot \sup_{i\geq 1}\mathbb{E}\left[\|\Gamma_{i,T}^{-1}\|^2 \cdot \|\Gamma_{i,T} - \Gamma_i\|^2 \cdot \|S_{i,T}\|^2\right]$$

$$\leq C\sup_{i\geq 1}\left(\mathbb{E}[\|\Gamma_{i,T}^{-1}\|^{2+\delta}]\right)^{\frac{2}{2+\delta}}\sup_{i\geq 1}\left(\mathbb{E}\left[\|\Gamma_{i,T} - \Gamma_i\|^{\frac{4(2+\delta)}{\delta}}\right]\right)^{\frac{\delta}{2(2+\delta)}}\sup_{i\geq 1}\left(\mathbb{E}\left[\|S_{i,T}\|^{\frac{4(2+\delta)}{\delta}}\right]\right)^{\frac{\delta}{2(2+\delta)}}. \tag{51}$$

现在考虑（51）右侧的各项，由于 $\Gamma_{i,T}$ 的一致可积性以及假设 4.4，我们有 $\sup_{i\geq 1}\mathbb{E}[\|\Gamma_{i,T}^{-1}\|^{2+\delta}] \leq C$，其中某个常数为 $C > 0$。此外，根据引理A.15，$\sup_{i\geq 1}(\mathbb{E}\|\Gamma_{i,T} - \Gamma_i\|^p)^{2/p} = O(T^{-1})$，且 $\sup_{i\geq 1}(\mathbb{E}\|S_{i,T}\|^p)^{2/p} = O(1)$。将这些结果代入（49）–（51），得到

$$\mathbb{E}\Big[\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Gamma_{i,T}^{-1} - \Gamma_i^{-1})S_{i,T}\Big\|^2\Big] \leq \sup_{i\geq 1}\mathbb{E}[\|\Gamma_i^{-1}(\Gamma_i - \Gamma_{i,T})\Gamma_{i,T}^{-1}S_{i,T}\|^2] = O\Big(\frac{1}{T}\Big). \tag{52}$$

由于 $T \to \infty$，利用 (48)，因此我们已经证明

$$\sqrt{nT}(\bar{\beta} - \beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma_i^{-1}S_{i,T} + o_p(1).$$

因此，对于我们的 渐近协方差 $V = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}$，接下来只需证明 t

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma_i^{-1}S_{i,T} = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\Gamma_i^{-1}X_i^{\mathsf{T}}\varepsilon_i = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\Gamma_i^{-1}x_{i,t}\varepsilon_{i,t} \xrightarrow{d} N(0,V).$$

任取一个 $v \in \mathbb{S}^{d-1}$，并令 $z_{i,t}(v) := v^{\mathsf{T}}\Gamma_i^{-1}x_{i,t}\varepsilon_{i,t}$，然后我们证明

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} z_{i,t}(v) \xrightarrow{d} N(0, v^{\mathsf{T}}Vv), \tag{53}$$

并通过 Cramér-Wold 得出结论。为此，令 $b_{i,T}(v) := \frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{i,t}(v)$，以及 $S_{n,T} := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} b_{i,T}(v)$。采用这一记号，(53) 变为 $S_{n,T} \xrightarrow{d} N(0, v^{\mathsf{T}}Vv)$，当 $(n,T) \to \infty$ 时成立。

我们将对数组 $\{\frac{1}{\sqrt{n}} b_{i,T}(v)\}$ 应用适用于三角阵的林德伯格-费勒中心极限定理，其中令 T 和 n 均趋于无穷大。具体而言，令 m 表示数组的行，并任取一个在两个坐标上均发散的序列 (n_m, T_m)。每一行中的元素将为 $\frac{1}{\sqrt{n_m}} b_{i,T_m}(v)$，使得行和为 $S_m(v) := \frac{1}{\sqrt{n_m}}\sum_{i=1}^{n_m} b_{i,T_m}(v) = S_{n_m,T_m}(v)$。由于 n_m 发散，因此这是一个三角阵。特别要注意，根据截面独立性这一假设，每一行中的所有元素彼此独立。

接下来，我们利用截面独立性、由外生性和零均值误差得到的 $S_m(v)$ 的零均值性质，以及两个过程的二阶平稳性，求出 $S_m(v)$的渐近方差。我们有，

$$\begin{aligned} V_m(v)^2 := \mathrm{Var}(S_m(v)) &= \frac{1}{n_m}\sum_{i=1}^{n_m}\mathrm{Var}(b_{i,T_m}(v)) = \frac{1}{n_m}\sum_{i=1}^{n_m}\mathrm{Var}\Big(\frac{1}{\sqrt{T_m}}\sum_{t=1}^{T_m} z_{i,t}(v)\Big) \\ &= \frac{1}{n_m}\sum_{i=1}^{n_m}\mathrm{Var}\Big(\frac{1}{\sqrt{T_m}}\sum_{t=1}^{T_m} v^{\mathsf{T}}\Gamma_i^{-1}x_{i,t}\varepsilon_{i,t}\Big) \\ &= v^{\mathsf{T}}\frac{1}{n_m T_m}\sum_{i=1}^{n_m}\Gamma_i^{-1}\sum_{t=1}^{T_m}\sum_{s=1}^{T_m}\mathbb{E}[x_{i,t}\varepsilon_{i,t}\varepsilon_{i,s}x_{i,s}^{\mathsf{T}}]\Gamma_i^{-1}v, \end{aligned} \tag{54}$$

且由于 $(x_{i,t}\varepsilon_{i,t})_{t\in\mathbb{Z}}$ 假设为弱平稳过程，且根据严格外生性，我们有

$$\sum_{t=1}^{T_m}\sum_{s=1}^{T_m}\mathbb{E}[x_{i,t}\varepsilon_{i,t}\varepsilon_{i,s}x_{i,s}^{\mathsf{T}}] = \sum_{t=1}^{T_m}\sum_{s=1}^{T_m}\mathbb{E}[x_{i,0}\varepsilon_{i,0}\varepsilon_{i,|t-s|}x_{i,|t-s|}^{\mathsf{T}}] = \sum_{|\ell|<T_m}(T_m - |\ell|)\mathbb{E}[x_{i,0}x_{i,\ell}^{\mathsf{T}}\varepsilon_{i,0}\varepsilon_{i,\ell}], \tag{55}$$

因此，我们可以将 (55) 应用于 (54)，得到

$$\begin{aligned} V_m(v)^2 &= v^{\mathsf{T}}\frac{1}{n_m T_m}\sum_{i=1}^{n_m}\Gamma_i^{-1}\sum_{|\ell|<T_m}(T_m - |\ell|)\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}]\Gamma_i^{-1}v \\ &= v^{\mathsf{T}}\frac{1}{n_m}\sum_{i=1}^{n_m}\Gamma_i^{-1}\sum_{\ell=-\infty}^{\infty}\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}]\Gamma_i^{-1}v + v^{\mathsf{T}}\frac{1}{n_m}\sum_{i=1}^{n_m}\Gamma_i^{-1}R_{i,T_m}\Gamma_i^{-1}v \\ &\to v^{\mathsf{T}}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}\right)v = v^{\mathsf{T}}Vv, \end{aligned}$$

其中我们定义了 $R_{i,\ T_m} = \sum_{|\ell|<T_m}(1-\frac{|\ell|}{T_m})\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}] - \sum_{\ell=-\infty}^{\infty}\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}]$ 和 $\Lambda_i := \sum_{\ell=-\infty}^{\infty}\mathbb{E}[x_{i,0}x_{i,\ell}^{\mathsf{T}}\varepsilon_{i,0}\varepsilon_{i,\ell}]$。特别是，最后一个极限使用了

$$\begin{aligned}\|R_{i,T_m}\| &= \Big\|\sum_{|\ell|<T_m}\Big(1-\frac{|\ell|}{T_m}\Big)\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}] - \sum_{\ell=-\infty}^{\infty}\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}]\Big\| \\ &\le \sum_{|\ell|>T_m}\|\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}]\| + \frac{1}{T_m}\sum_{|\ell|<T_m}|\ell|\|\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}]\| = o(1).\end{aligned}$$

根据我们的混合假设，当 $T_m\to\infty$ 时，这个等式最后一步对 i 一致成立。特别地，由引理 A.10 可知，零均值过程 $(x_{i,\ t}\varepsilon_{i,\ t})$ 是强混合的，其强混合系数以 $\alpha_{score}(k) = O(k^{-3-q})$ 的速度衰减。结合引理 A.11 可得 $\sum_{|\ell|=-\infty}^{\infty}|\ell|\|\mathbb{E}[x_{i,l}x_{i,0}^{\mathsf{T}}\varepsilon_{i,l}\varepsilon_{i,0}]\| < \infty$，因此第二项在 $T_m\to\infty$ 时趋于零。同理，$\sum_{\ell=-\infty}^{\infty}\|\mathbb{E}[x_{i,l}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}]\| < \infty$ 立即意味着 $\sum_{|\ell|>T_m}\|\mathbb{E}[x_{i,\ell}x_{i,0}^{\mathsf{T}}\varepsilon_{i,\ell}\varepsilon_{i,0}]\|\to 0$ 在 $T_m\to\infty$ 时成立。

接下来验证林德伯格条件，即对于所有 $\epsilon>0$,

$$\lim_{m\to\infty}\sum_{i=1}^{n_m}\frac{1}{n_m V_m(v)^2}\mathbb{E}\left[b_{i,T_m}^2\mathbb{1}\{|b_{i,T_m}(v)|\ge\sqrt{n_m}\epsilon V_m(v)\}\right] = 0. \tag{56}$$

利用对于任意 $0<\delta<2$，我们有 $|x|^2\mathbb{1}\{|x|>a\} \le \frac{|x|^{2+\delta}}{a^{\delta}}$ 得到

$$\mathbb{E}\left[b_{i,T_m}^2\mathbb{1}\{|b_{i,T_m}(v)|\ge\sqrt{n_m}\epsilon V_m(v)\}\right] \le \frac{\mathbb{E}\left[|b_{i,T_m}|^{2+\delta}\right]}{(\sqrt{n_m}\epsilon V_m(v))^{\delta}}. \tag{57}$$

此外，由引理 A.12 可得

$$\sup_{i,T}\mathbb{E}[|b_{i,T}(v)|^{2+\delta}] = \sup_{i,T}\mathbb{E}\Big[\frac{1}{T^{1+\delta/2}}\Big|\sum_{t=1}^{T}v^{\mathsf{T}}\Gamma_i^{-1}x_{i,t}\varepsilon_{i,t}\Big|^{2+\delta}\Big] = \sup_{i,T}\frac{1}{T^{1+\delta/2}}O(T^{1+\delta/2}) = O(1). \tag{58}$$

引理 A.12 的条件之所以满足，是因为根据引理 A.10，零均值过程 $(x_{i,\ t}\varepsilon_{i,\ t})$ 是强混合的，其强混合系数 $\alpha_{score_i}(k)$ 满足

$$\sup_{i\ge1}\sum_{k=1}^{\infty}(k+1)^{c-2}\alpha_{score_i}(k)^{\frac{\epsilon}{c+\epsilon}} < \infty, \tag{59}$$

对某个 $\epsilon>0$，$c\in2\mathbb{N}$，$c>2+\delta$。特别地，由于 $\delta<2$，若取 $c=4$，则有

$$\sum_{k=1}^{\infty}(k+1)^2\alpha_{score_i}(k)^{\frac{\epsilon}{4+\epsilon}} \le C\sum_{k=1}^{\infty}k^2k^{(-3-q)\frac{\epsilon}{4+\epsilon}},$$

且对于任意 $\epsilon>\frac{12}{q}$，该值都是有限的。

因此，由（56）、（57）和（58）可得

$$\sum_{i=1}^{n_m}\frac{1}{n_m V_m(v)^2}\mathbb{E}\left[b_{i,T_m}^2\mathbb{1}\{|b_{i,T_m}(v)|\ge\sqrt{n_m}\epsilon V_m(v)\}\right] = \frac{O(1)}{n_m^{\delta/2}\epsilon^{\delta}V_m(v)^{2+\delta}}\to 0, \tag{60}$$

由于 $n_m\to\infty$, $\epsilon = O(1)$，且 $V_m(v)^2\to v^{\mathsf{T}}(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})v > 0$。因此，我们已验证林德伯格条件，并得出 $\frac{S_m(v)}{V_m(v)}\xrightarrow{d}N(0,1)$。由于这对

每个满足 $n_m \to \infty, T_m \to \infty$ 的序列 $(n_m, T_m)_{m\geq 1}$ 都成立，因此，利用斯卢茨基定理，

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{i,t}(v) \xrightarrow{d} N(0, v^{\mathsf{T}} V v),$$

当 $(n, T) \to \infty$ 时。由克拉默-沃尔德定理，最终可得

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \Gamma_i^{-1} X_i^{\mathsf{T}} \varepsilon_i \xrightarrow{d} N(0, V),$$

当 $(n, T) \to \infty$ 时，因此，根据斯卢茨基定理

$$\sqrt{nT}(\bar{\beta} - \beta) \xrightarrow{d} N(0, V) \quad \text{and} \quad \sqrt{nT}(\hat{\beta}_{\mathsf{DP}} - \beta) \xrightarrow{d} N(0, V).$$

□

C.4. 定理 4.8 的证明。 我们首先证明算法 4 的隐私和效用保证；算法 4 是第 4.3 节的核心。具体而言，定理 C.1 将使我们能够推导出定理 4.8 的保证。

算法 4 差分隐私处理的均值估计 DPTreatMean D z n B R $(\ _i, \ _i)_{i=1}, \mu, \ , \ , \xi$

Input: Data $(D_i, z_i)_{i=1}^n$, privacy parameter μ, radius B, max rounds R, failure probability ξ
Noisy size: $\tilde{z} \leftarrow \sum_{i=1}^n z_i + \frac{2}{\mu} Z_0, \quad Z_0 \sim \mathcal{N}(0,1)$
Set: $\tau \leftarrow \max\Big\{\tilde{z} - \frac{2\sqrt{R}}{\mu}\sqrt{2\log\frac{4R}{\xi}} - \frac{2}{\mu}\sqrt{2\log\frac{8}{\xi}}, 1\Big\}$, $n_{LB} \leftarrow \max\Big\{\tau - \frac{2\sqrt{R}}{\mu}\sqrt{2\log\frac{4R}{\xi}}, 1\Big\}$
Init: $\hat{m}_{r-2} \leftarrow 0$, $\hat{m}_{r-1} \leftarrow 0$ $r \leftarrow 0$, $r^* \leftarrow -1$
while $r \leq R$ and $r^* = -1$ **do**
  $\omega_i \leftarrow \mathbb{1}\{\|D_i - \hat{m}_{r-1}\| < B z_i / 2^r\}, \quad \omega \leftarrow \sum_{i=1}^n z_i \omega_i + \frac{2\sqrt{R}}{\mu} Z_r, \quad Z_r \sim \mathcal{N}(0,1)$
  if $\omega < \tau$ **then**
    $S \leftarrow \{i : \|D_i - \hat{m}_{r-2}\| < B z_i / 2^{r-1}\}, \quad n_{\tilde{S}} \leftarrow \max(|S|, n_{LB}), \quad \mathcal{C} \leftarrow \sqrt{3 - \frac{2r}{R}}$,
    $\hat{m} \leftarrow \hat{m}_{r-2} + \frac{1}{n_{\tilde{S}}} \sum_{i\in S}(D_i - \hat{m}_{r-2}) + \frac{4B}{2^{r-1}\mathcal{C}\mu n_{LB}} N_*, \quad N_* \sim \mathcal{N}(0, I)$
    $r^* \leftarrow r - 1, \quad B^* \leftarrow \frac{4B}{2^{r^*}\mathcal{C}\mu n_{LB}}$
  else if $r = R$ **then**
    $S \leftarrow \{i : \|D_i - \hat{m}_{r-1}\| < B z_i / 2^R\}, \quad n_{\tilde{S}} \leftarrow \max(|S|, n_{LB})$
    $\hat{m} \leftarrow \hat{m}_{r-1} + \frac{1}{n_{\tilde{S}}} \sum_{i\in S}(D_i - \hat{m}_{r-1}) + \frac{4B}{2^R \mu n_{LB}} N_*, \quad N_* \sim N(0, I)$
    $r^* \leftarrow R, \quad B^* \leftarrow \frac{4B}{2^{r^*}\mu n_{LB}}$,
  else
    $S \leftarrow \{i : \|D_i - \hat{m}_{r-1}\| < B z_i / 2^r\}, \quad n_{\tilde{S}} \leftarrow \max(|S|, n_{LB})$
    $\hat{m}_r \leftarrow \hat{m}_{r-1} + \frac{1}{n_{\tilde{S}}} \sum_{i\in S}(D_i - \hat{m}_{r-1}) + \frac{4B\sqrt{R}}{2^r \mu n_{LB}} N_r, \quad N_r \sim N(0, I)$
    $r \leftarrow r + 1$
  end if
end while
Output: $(\hat{m}, \hat{m}_{r^*-1}, r^*, n_{LB}, B^*)$

定理 C.1。 算法 *4* 满足 μ-高斯差分隐私。此外，假设存在 B、K_1、$K_2 > 0^{-}$，使得 最大值 $_i \|z_i D_i\| \le B$、$\|D_{z=1} - m\| \le K_1$ 以及 最大值$_{i:z_i=1} \|D_i - m\| \le K_2$，并假设 $\sum_{i=1}^n z_i =: \bar{z} \asymp n$。如果 $R \ge C$对数 $\frac{B}{K_2}$且 $n \ge C'(\frac{1}{\mu}\sqrt{R}($对数 $\frac{R}{\xi} + d))$ 对于通用常数 C、$C' > 0$成立，则算法 *4* 以至少 $1 - \xi$ 的概率满足

$$\|\hat{m}_{\mathsf{DP}} - m\| = O\Big(K_1 + \frac{K_2}{n\mu}\sqrt{R\log\frac{R}{\xi}} + \frac{K_2}{n\mu}\sqrt{d + \log\frac{1}{\xi}}\Big).$$

Proof. 为简单起见，我们有时会使用标签 $\mu_1 = \mu_2 = \frac{\mu}{\sqrt{4R}}$ 和 $\mu_3 = \frac{}{\mu_2}$。

隐私证明与命题 B.1 的证明类似。事实上，最终输出仅通过统计量 $\sum_{i=1}^n z_i$、$\sum_{i=1}^n z_i\omega_i$ 和 $\frac{1}{n_{\tilde{S}}}\sum_{i\in S}$（$D_i - \hat{}m_{r-j}$）依赖于数据，因此，只要证明这些统计量具有足够的隐私性，我们的隐私结果便可通过组合得出（命题 2.7）。

首先，由于 $z_i \in \{0, 1\}$，显然 $\sum_{i=1}^n z_i$ 的敏感度不超过 1，因此统计量 $\bar{z}$ 满足 μ_3-高斯差分隐私。统计量 $\sum_{i=1}^n z_i\omega_i$ 也同样如此，因为 $z_i\omega_i \in \{0, 1\}$，所以第 7 行的每个实例都满足 μ_2-高斯差分隐私。

最后，我们考虑 $\frac{1}{n_{\tilde{S}}}\sum_{i\in S} z_i(D_i - \hat{}m_{r-j})$,的敏感度，其中 $n_{\tilde{S}} = \max(|S|, n_{LB})$，且 $j = 1$ 出现在第 13 行和第 17 行，$j = 2$ 出现在第 9 行。令 $(x_i, z_i)_{i=1}^n$ 和 $(x_i', z_i')_{i=1}^n$ 为两个相邻数据集；不失一般性，假设这两个数据集只有第一项不同。为了推导该统计量的敏感度，我们必须考虑两种情况。首先考虑 $z_1 = z_1'$ 的情况。在这种情况下，分析与算法 1 中的分析完全相同，因此我们立即得知，该情况下的敏感度不超过 $\frac{2B}{2^{r-(j-1)}n_{LB}}$。因此，我们深入考察 $z_1' = z_1 + 1^{-}$的情况（敏感度具有对称性，因此无需考虑 $z_1' = z_1 - 1$）。如果 $\|D_1' - \hat{}m_{r-j}\| > \frac{B}{2^r},$，那么 $S = S', n_{\tilde{S}} = n_{\tilde{S}}'$，因此差值为 0。然而，如果 $\|D_1' - \hat{}m_{r-j}^{-}\| \le \frac{B}{2^r},$，那么 $S \ne S'$。现在有两种情况需要考虑。首先，如果 $|S'| \le n_{LB}$，那么我们有 $n_{\tilde{S}} = n_{\tilde{S}}'$，从而可以将敏感度界定为

$$\Big\|\frac{1}{n_{\tilde{S}}}\sum_{i\in S} z_i(D_i - \hat{m}_{r-j}) - \frac{1}{n_{\tilde{S}}'}\sum_{i\in S'} z_i'(D_i' - \hat{m}_{r-j})\Big\| = \Big\|\frac{1}{n_{\tilde{S}}}(D_1' - \hat{m}_{r-j})\Big\| \le \frac{B}{2^{r-(j-1)}n_{LB}}.$$

否则，如果 $|S'| > n_{LB}$，则 $n_{\tilde{S}} \ne n_{\tilde{S}}'$。在这种情况下，

$$\begin{aligned}
&\Big\|\frac{\sum_{i\in S} z_i(D_i - \hat{m}_{r-j})}{n_{\tilde{S}}} - \frac{\sum_{i\in S'} z_i'(D_i' - \hat{m}_{r-j})}{n_{\tilde{S}}'}\Big\| \\
&\quad = \Big\|\frac{(n_{\tilde{S}} + 1)\sum_{i\in S} z_i(D_i - \hat{m}_{r-j}) - n_{\tilde{S}}\sum_{i\in S'} z_i'(D_i' - \hat{m}_{r-j})}{n_{\tilde{S}}(n_{\tilde{S}} + 1)}\Big\| \\
&\quad \le \Big\|\frac{\sum_{i\in S} z_i(D_i - \hat{m}_{r-j}) - n_{\tilde{S}}(D_1' - \hat{m}_{r-j})}{n_{\tilde{S}}(n_{\tilde{S}} + 1)}\Big\|.
\end{aligned}$$

为完成证明，我们注意到

$$\left\|\frac{\sum_{i\in S} z_i(D_i-\hat{m}_{r-j})-n_{\tilde{S}}(D_1'-\hat{m}_{r-j})}{n_{\tilde{S}}(n_{\tilde{S}}+1)}\right\| \le \left\|\frac{n_{\tilde{S}}B}{n_{\tilde{S}}(n_{\tilde{S}}+1)2^{r-(j-1)}}\right\| + \left\|\frac{n_{\tilde{S}}B}{n_{\tilde{S}}(n_{\tilde{S}}+1)2^{r-(j-1)}}\right\|$$
$$\le \frac{2B}{(n_{\tilde{S}}+1)2^{r-(j-1)}} \le \frac{2B}{(n_{LB}+1)2^{r-(j-1)}} \le \frac{2B}{2^{r-(j-1)}n_{LB}},$$

由此可得该量的敏感度，因为其他所有情形都可类似推得。

现在，若 $r^* = R$，则最终输出是通过调用一次第 2 行、调用一次第 13 行、调用 R 次第 7 行以及调用 R 次第 17 行，由数据得到的函数；我们已证明这些调用分别满足 μ_3、 μ_3、 μ_3、 μ_2 和 μ_1-高斯差分隐私。因此可得，$(\hat{m},\ r^*)$ 满足 $\sqrt{R(\mu_1^2+\mu_2^2)+2\mu_3^2}$-uGDP。然而，若算法以 $r^* < R$ 终止，则算法的输出是通过调用一次第 2 行、以隐私参数 μ_3C 调用一次第 9 行、调用 r^*+1 次第 7 行以及调用 r^*+1 次第 17 行，由数据得到的函数。因此在这种情况下，$(\hat{m},\ r^*)$ 满足 $\sqrt{(1+C^2)\mu_3^2+(r^*+1)(\mu_1^2+\mu_2^2)}$-uGDP；并且不难验证，对于我们选择的 C，这等于 $\sqrt{R(\mu_1^2+\mu_2^2)+2\mu_3^2}$。

现在我们继续证明效用保证。该证明与基础算法的证明类似，只需稍作调整以处理纳入变量z_i的隐私化。首先，我们定义一个高概率事件，在该事件下，定理 3.1 中的分析可以很好地应用于算法 4。令

$$\mathcal{E}_0 = \left\{|Z_0| \le \sqrt{2\log\frac{8}{\xi}}\right\},$$

根据 引理 A.2，我们知道其以至少 $1-\xi/4$ 的概率成立。此外 ，

$$\mathcal{E}_1 = \left\{\max_{r\in[R]}|Z_r| \le \sqrt{2\log\frac{4R}{\xi}}\right\},$$

根据引理 A.2，我们还知道其以至少 $1-\frac{\xi}{2}$ 的概率成立。此外， –

$$\mathcal{E}_2 = \left\{\max_{r\in 0,\dots,R-1}\|N_r\| \le \sqrt{d}+\sqrt{2\log\frac{8(R+1)}{\xi}}\right\},$$

根据引理 A.3，我们知道其以至少 $1-\frac{\xi}{8}$ 的概率成立，并且 –

$$\mathcal{E}_3 = \left\{\|N_*\| \le \sqrt{d}+\sqrt{2\log\frac{8}{\xi}}\right\},$$

其以至少 $1-\frac{\xi}{8}$ 的概率成立。因此，事件

$$\mathcal{E} := \mathcal{E}_0\cap\mathcal{E}_1\cap\mathcal{E}_2\cap\mathcal{E}_3 \tag{61}$$

hold以至少 $1-\xi$ 的概率成立，并且我们在该事件下分析算法 4 .

令 $\mathcal{Z} := \{i : z_i = 1\}$，且 $\bar{D}_{\mathcal{Z}} := \frac{1}{\tilde{z}}\sum_{i\in\mathcal{Z}} D_i$。总体而言，算法 4 与将算法 3 应用于 $(D_i)_{i\in\mathcal{Z}}$ 相同，唯一的区别是用带噪样本量 $\tilde{z}$ 替换真实样本量$\bar{z}$ ，并相应地改变 τ 和 n_{LB}。请注意，只有通过改变 τ 和 n_{LB} ，算法 4 的输出空间才会与算法 3 在输入为 $(:=)i : z_i = 1$ 时的输出空间不同。

不过，在新的高概率事件 E 上，我们仍有：若 $\sum_{i=1}^{n}\omega_i=\bar{z}$，则 $\omega\geq\tau$。然后，如式 (19) 所示，对于 $i\in\mathcal{Z}$，我们得到：

$$\|D_i-\hat{m}_{r-1}\|\leq\|D_i-\bar{D}_{\mathcal{Z}}\|+\frac{4B\sqrt{R}}{2^{r-1}\mu n_{LB}}\|N_{r-1}\|,$$

因此，按照 (19) 附近的论证，我们可以得出，只要 $2K_2\leq B/2^{r+1}$ 和 $\bar{z}\geq C\frac{1}{\mu}\sqrt{R(\log\frac{R}{\xi}+d)}$，算法就不会终止。为此，记将算法 1 应用于 $(D_i)_{i\in\mathcal{Z}}$ 时 n_{LB} 的取值为 n_{LB_1}，将算法 4 应用于 $(D_i,\ z_i)_{i=1}^{n}$ 时得到的 n_{LB} 的取值为 n_{LB_2}。那么，

$$n_{LB_2}=n_{LB_1}+\frac{1}{\mu_3}\Big(Z_0-\sqrt{2\log\frac{8}{\xi}}\Big),\tag{62}$$

其中，我们假设 $\sum_{i=1}^{n}z_i$ 足够大，使得在我们的高概率分析中这些量都大于 1。在我们的假设下，在 E 上有

$$\Big|\frac{1}{\mu_3}\Big(Z_0-\sqrt{2\log\frac{8}{\xi}}\Big)\Big|\leq\frac{2}{\mu_3}\sqrt{2\log\frac{8}{\xi}},$$

因此，在这一高概率事件上，对 $\bar{z}$ 的要求与算法 3 的设定中 (20) 的量级相同。由此可得，在事件 E 上，只要 $\bar{z}\geq C\frac{1}{\mu}\sqrt{R(\log\frac{R}{\xi}+d)}$，算法 4 就能实现 $B/2^{r^*}\leq 2K_2$。

最终的误差界可由定理 3.1 中第 (22)-(18) 行附近的效用论证得出，只需注意到，正如前面所论证的，$n_{LB_2}\asymp n_{LB_1}\geq c\bar{z}\asymp n$ 对于某个常数 $c>0$成立。因此，

$$\|\hat{m}-m\|=O\Big(K_1+\frac{K_2}{n\mu}\sqrt{R\log\frac{R}{\xi}}+\frac{K_2}{n\mu}\sqrt{d+\log\frac{1}{\xi}}\Big).$$

□

现在我们可以给出定理 4.8 的证明了。

证明。 根据定理 C.1，令 K_1 和 K_2 以与定理 4.6 的证明相同的方式获得（即使用引理 4.3），我们知道，以至少 $1-\xi/2$ 的概率，

$$\|\hat{\beta}_{1,\mathsf{DP}}-(\beta+\theta)\|=O\Big(\sigma_\varepsilon\sqrt{\frac{M_1(d+\log\frac{1}{\xi})}{nT}}+\sigma_\varepsilon\max_{i:z_i=1}\|\Gamma_i^{-\frac{1}{2}}\|\sqrt{\frac{(R\log\frac{R}{\xi}+d)(d+\log\frac{n}{\xi})}{n^2\mu^2T}}\Big),$$

因为 $\bar{z}=\Omega(n)$，并且类似地，以至少 $1-\xi/2$ 的概率，

$$\|\hat{\beta}_{0,\mathsf{DP}}-(\beta+\theta)\|=O\Big(\sigma_\varepsilon\sqrt{\frac{M_0(d+\log\frac{1}{\xi})}{nT}}+\sigma_\varepsilon\max_{i:z_i=0}\|\Gamma_i^{-\frac{1}{2}}\|\sqrt{\frac{(R\log\frac{R}{\xi}+d)(d+\log\frac{n}{\xi})}{n^2\mu^2T}}\Big),$$

又因为 $n_0 \asymp n$。因此，以至少 $1-\xi$ 的概率，

$$\|\hat{\beta}_{1,\mathsf{DP}} - \hat{\beta}_{0,\mathsf{DP}} - \theta\| = \|\hat{\beta}_{1,\mathsf{DP}} - (\beta+\theta) - (\hat{\beta}_{0,\mathsf{DP}} - \beta)\| \le \|\hat{\beta}_{1,\mathsf{DP}} - (\beta+\theta)\| + \|\hat{\beta}_{0,\mathsf{DP}} - \beta\|$$
$$= O\Big(\sigma_\varepsilon \sqrt{\frac{(M_0+M_1)(d+\log\frac{1}{\xi})}{nT}} + \sigma_\varepsilon \max_{i\in[n]} \|\Gamma_i^{-\frac{1}{2}}\| \sqrt{\frac{(R\log\frac{R}{\xi}+d)(d+\log\frac{n}{\xi})}{n^2\mu^2 T}}\Big).$$

此外，通过完全遵循定理 4.6 证明中的步骤，可以得出，在我们的增长条件下，我们有

$$\sqrt{pnT}\left(\hat{\beta}_{1,\mathsf{DP}} - (\beta+\theta)\right) \xrightarrow{d} N\Big(0, (\lim_{n\to\infty} \frac{1}{pn} \sum_{i\in G_1} \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})\Big),$$

同样，

$$\sqrt{(1-p)nT}\left(\hat{\beta}_{0,\mathsf{DP}} - \beta\right) \xrightarrow{d} N\Big(0, (\lim_{n\to\infty} \frac{1}{(1-p)n} \sum_{i\in G_2} \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})\Big),$$

由于二者相互独立，因此可以得出，

$$\sqrt{nT}(\hat{\beta}_{1,\mathsf{DP}} - \hat{\beta}_{0,\mathsf{DP}} - \theta) \xrightarrow{d} N\Big(0, \frac{1}{p}(\lim_{n\to\infty} \frac{1}{pn} \sum_{i\in G_1} \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}) + \frac{1}{1-p}(\lim_{n\to\infty} \frac{1}{(1-p)n} \sum_{i\in G_2} \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})\Big).$$

隐私性源于定理 C.1 的隐私证明，并注意到该定理陈述中的隐私参数已按 $\frac{1}{\sqrt{2}}$ 重新调整，以考虑在两组查询上的组合。

□

Appendix D. Proofs of Section 5

D.1. 定理 5.1 的证明。

ˆ证明。 我们首先证明隐私结果。量 (β_{DP}、ˆm_{r^*-1}， r^*， B^*) 是通过 μ_{est}-uGDP 估计的。由此可知，若将 (β_{DP}、ˆm_{r^*-1}， r^*， B^*) 视为固定的非敏感量时，所得协方差统计量 $V(\beta_{\mathsf{DP}})_{\mathsf{DP}}$ˆ 满足 μ_{var}-uGDP，那么根据高斯差分隐私的组合性质（命题 2.7），组合输出空间 (ˆβ_{DP}、ˆV(ˆβ_{DP})$_{\mathsf{DP}}$) 满足 $\sqrt{\mu_{\mathsf{est}}^2 + \mu_{\mathsf{var}}^2}$-uGDP。

现在我们证明算法 2 中计算得到）在将（β_{DP}，ˆm_{r^*-1}， r^*， B^*）视为任意固定量时，满足 μ_{var}-uGDP。回顾算法 2 第 2 行中定义的 S和 $n_{\tilde{S}}$。考虑两个相邻数据集 D 和 D′，不失一般性地，我们改变第一个用户的数据。于是，协方差构造的弗罗贝尼乌斯范数敏感度是下式的最小上界

$$\Delta := \Big\| \frac{1}{n_{\tilde{S}}^2} \sum_{i\in S} (\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})^{\mathsf{T}} - \frac{1}{(n'_{\tilde{S}})^2} \sum_{i\in S'} (\hat{\beta}'_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}'_i - \hat{\beta}_{\mathsf{DP}})^{\mathsf{T}} \Big\|_F.$$

我们希望找到敏感度 Δ 的一个上界。首先考虑$1 \in S$ 和 $1 \in S'$的情况。于是 $S = S'$ 和 $n_{\tilde{S}} = n'_{\tilde{S}}$，因此

$$\begin{aligned}\Delta &= \Big\| \frac{1}{n_{\tilde{S}}^2} \sum_{i \in S} \big((\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} - (\hat{\beta}'_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}'_i - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} \big) \Big\|_F \\ &= \Big\| \frac{1}{n_{\tilde{S}}^2} \big((\hat{\beta}_1 - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}_1 - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} - (\hat{\beta}'_1 - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}'_1 - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} \big) \Big\|_F \\ &\le \frac{1}{n_{\tilde{S}}^2} \sqrt{\|\hat{\beta}_1 - \hat{\beta}_{\mathsf{DP}}\|^4 + \|\hat{\beta}'_1 - \hat{\beta}_{\mathsf{DP}}\|^4} \le \frac{1}{n_{\tilde{S}}^2} \sqrt{\kappa^4 + \kappa^4} \le \frac{\sqrt{2}\kappa^2}{n_{LB}^2},\end{aligned}$$

其中我们利用了 $1 \in S$，S'根据 S，S' 的定义可推出 $\|\beta_1 - \beta_{\mathsf{DP}}\| \le \kappa$，因为$\|\hat{\beta}_1 - \hat{\beta}_{\mathsf{DP}}\| \le \|\hat{\beta}_1 - \hat{m}_{r^*-1}\| + \|\hat{m}_{r^*-1} - \hat{\beta}_{\mathsf{DP}}\| \le \frac{B}{2^{r^*}} + \delta_{\mathsf{DP}} = \kappa$，对于 $\hat{\beta}'_1$ 也同理。

另一方面，如果 $1 \notin S$ 且 $1 \in S'$，那么必然有 $S' = S \cup \{1\}$，因为这两个数据集是相邻的，因此 $|S'| = 1 + |S|$。由此可将敏感度 Δ 界定为

$$\begin{aligned}\Delta &= \Big\| \frac{1}{n_{\tilde{S}}^2} \sum_{i \in S} (\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} - \frac{1}{(n'_{\tilde{S}})^2} \sum_{i \in S'} (\hat{\beta}'_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}'_i - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} \Big\|_F \\ &= \Big\| \Big(\frac{1}{n_{\tilde{S}}^2} - \frac{1}{(n'_{\tilde{S}})^2} \Big) \sum_{i \in S} (\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} - \frac{1}{(n'_{\tilde{S}})^2} (\hat{\beta}'_1 - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}'_1 - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} \Big\|_F \\ &\le \Big\| \frac{2n_{\tilde{S}} + 1}{n_{\tilde{S}}^2 (n'_{\tilde{S}})^2} \sum_{i \in S} (\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} \Big\|_F + \Big\| \frac{1}{(n'_{\tilde{S}})^2} (\hat{\beta}'_1 - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}'_1 - \hat{\beta}_{\mathsf{DP}})^\mathsf{T} \Big\|_F \\ &\le \frac{2n_{\tilde{S}} + 1}{n_{\tilde{S}}^2 (n'_{\tilde{S}})^2} \sum_{i \in S} \|\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}}\|^2 + \frac{1}{(n'_{\tilde{S}})^2} \|\hat{\beta}'_1 - \hat{\beta}_{\mathsf{DP}}\|^2 \\ &\le \frac{(2n_{\tilde{S}} + 1)|S|\kappa^2}{n_{\tilde{S}}^2 (n'_{\tilde{S}})^2} + \frac{\kappa^2}{(n'_{\tilde{S}})^2} \le \frac{n_{\tilde{S}}(2n_{\tilde{S}} + 1)\kappa^2}{n_{\tilde{S}}^4} + \frac{\kappa^2}{n_{LB}^2} \le \frac{3n_{\tilde{S}}^2 \kappa^2}{n_{\tilde{S}}^4} + \frac{\kappa^2}{n_{LB}^2} \le \frac{4\kappa^2}{n_{LB}^2}. \qquad (63)\end{aligned}$$

根据对称性，当 $1 \in S$ 且 $1 \notin S'$ 时，敏感度相同。最后，如果 $1 \notin S$，S'，则敏感度显然为零。因此，无论哪种情况，敏感度均不超过 $\frac{4\kappa^2}{n_{LB}^2}$。

接下来，我们将论证，为了保证协方差估计量满足μ-高斯差分隐私，只需向对角线估计值添加 $N(0, \frac{16\kappa^4}{\mu^2 n_{LB}^4})$，并向非对角线估计值添加 $N(0, \frac{8\kappa^4}{\mu^2 n_{LB}^4})$。为此，考虑由下式给出的双射线性映射 $\mathsf{vec} : \mathbb{M}_{symm}^{d \times d} \mapsto \mathbb{R}^{d(d+1)/2}$

$$\mathrm{vec}(M) = (M_{11}, M_{22}, \ldots, M_{dd}, \sqrt{2}M_{12}, \ldots, \sqrt{2}M_{(d-1)d}),$$

并且请注意，对于任意$M \in \mathbb{M}_{symm}^{d \times d}$，该映射在如下意义下保持距离：$\|M\|_F = \|\mathsf{vec}(M)\|_2$。因此，使用 vec 将样本协方差矩阵映射到 $\mathbb{R}^{d(d+1)/2}$后，其 L_2-敏感度也为 $\frac{4\kappa^2}{n_{LB}^2}$。根据命题 2.5，若 $Z \sim N(0, I_{d(d+1)/2})$，则

$$\mathrm{vec}(M) + \frac{4\kappa^2}{n_{LB}^2 \mu_{\mathsf{var}}} Z \qquad (64)$$

满足 μ_{var}-高斯差分隐私。现在，为了还原对称矩阵，我们执行

$$\mathrm{vec}^{-1}\Big(\mathrm{vec}(M)+\frac{4\kappa^2}{n_{LB}^2\mu_{\mathsf{var}}}Z\Big)=M+\frac{4\kappa^2}{n_{LB}^2\mu_{\mathsf{var}}}\mathrm{vec}^{-1}(Z),$$

且 $\mathrm{vec}^{-1}(Z)$ 是将 Z 映射回实对称 $d\times d$ 矩阵并将非对角元素除以 $\sqrt{2}$ 的矩阵。根据后处理（命题 2.6），此输出满足 μ_{var}-高斯差分隐私。因此，根据组合，$(\hat{\beta}_{\mathsf{DP}},\ \hat{V}(\hat{\beta}_{\mathsf{DP}})_{\mathsf{DP}})$ 满足 $\sqrt{\mu_{\mathsf{est}}^2+\mu_{\mathsf{var}}^2}$-高斯差分隐私。

我们继续进行该算法的收敛分析。我们的分析将在事件 $\mathrm{E}=\mathrm{E}_0\cap\mathrm{E}_1\cap\mathrm{E}_2$, 上进行，其中 E_1 是引理 4.3 的高概率界成立的事件（其成立概率至少为 $1-\xi$），E_2 是在定理 4.6 的条件下将方程 (15)-(17) 应用于 $(\beta_i)_{i\in[n]}$ 时这些方程成立的事件；根据这些方程相关讨论，该事件的成立概率也至少为 $1-\xi$，以及事件

$$\tilde{\mathcal{E}}_0:=\Big\{\|W\|=O\Big(\frac{\kappa^2}{n_{LB}^2\mu_{\mathsf{var}}}\sqrt{d+\log\frac{1}{\xi}}\Big)\Big\}, \tag{65}$$

根据引理 A.5，该结论以至少 $1-\xi$ 的概率成立。因此，由并集界可知，E 以至少 $1-3\xi$ 的概率成立。因此，由于根据 (5) 我们有 $\xi=o(1)$，该事件发生的概率将趋于 1。

根据算法 2 的第 3 行，我们有

$$nT\hat{V}(\hat{\beta}_{\mathsf{DP}})_{\mathsf{DP}}=\frac{nT}{n_{\tilde{S}}^2}\sum_{i\in S}(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})^{\mathsf{T}}+nT(B^*)^2I+nTW, \tag{66}$$

其中，对角元素为 $W_{i,\,i}\overset{iid}{\sim}N(0,\ \frac{16\kappa^4}{n_{LB}^4\mu_{\mathsf{var}}^2})$，非对角元素为 $W_{i,\,j}\overset{iid}{\sim}N(0,\ \frac{8\kappa^4}{n_{LB}^4\mu_{\mathsf{var}}^2})$，对于 $j>i$，有 $W_{i,\,j}=W_{j,\,i}$，对于 $j<i$。接下来，我们将证明 (66) 右侧的第二项和第三项为 $o(1)$。

我们先分析 (66) 中的项 nTW。根据算法 2 的第 2 行，我们有

$$\kappa=\frac{B}{2^{r^*}}+\delta_{\mathsf{DP}}=\frac{B}{2^{r^*}}+\|\hat{m}_{r^*-1}-\hat{\beta}_{\mathsf{DP}}\|, \tag{67}$$

并且根据引理 B.2 的性质 2，$\frac{B}{2^{r^*}}<2K_2$ on $\tilde{\mathrm{E}}$。此外，由算法 1 第 8 行，并利用 S 的定义，

$$\begin{aligned}\delta_{\mathsf{DP}}=\|\hat{\beta}_{\mathsf{DP}}-\hat{m}_{r^*-1}\|&=\Big\|\frac{1}{n_{\tilde{S}}}\sum_{i\in S}(\hat{\beta}_i-\hat{m}_{r^*-1})+\frac{2\sqrt{2}B}{2^{r^*}\tilde{\mathcal{C}}\mu_{\mathsf{est}}n_{LB}}N_*\Big\|\\&\le\frac{1}{|S|}\sum_{i\in S}\frac{B}{2^{r^*}}+\frac{2\sqrt{2}B}{2^{r^*}\tilde{\mathcal{C}}\mu_{\mathsf{est}}n_{LB}}\|N_*\|\\&\le\frac{B}{2^{r^*}}+O\Big(\frac{B\sqrt{d+\log\frac{1}{\xi}}}{2^{r^*}\mu_{\mathsf{est}}n_{LB}}\Big)=O(K_2)+O\Big(K_2\frac{\sqrt{d+\log\frac{1}{\xi}}}{\mu_{\mathsf{est}}n_{LB}}\Big)=O(K_2),\end{aligned}$$

如果 $n \geq \frac{C}{\mu_{\mathsf{est}}}\sqrt{R(d+\log\frac{R}{\xi})}$)，由于此时根据引理 B.2 的性质 1，我们有 $n_{LB} = \Omega(n)$，且 ~~，其中我们使用了 E 的 $\|N_*\|$ 上界 (17)。因此，在 E 上有 $\kappa = O(K_2)$。由此，

$$\frac{4\kappa^2}{n_{LB}^2\mu_{\mathsf{var}}} = O\Big(\frac{K_2^2}{n^2\mu_{\mathsf{var}}}\Big) = O\Big(\sigma_\varepsilon^2 \max_{i\in[n]}\|\Gamma_i^{-\frac{1}{2}}\|^2 \frac{1}{Tn^2\mu_{\mathsf{var}}}\big(d+\log\frac{n}{\xi}\big)\Big),$$

~其中第二个等式由引理 4.3 得出。因此，我们可以将 E_0 与 (65) 及 (5) 结合使用，从而以概率 $1-\xi$ 得到，

$$\|nTW\| = O\Big(\sigma_\varepsilon^2 \max_{i\in[n]}\|\Gamma_i^{-\frac{1}{2}}\|^2 \frac{1}{n\mu_{\mathsf{var}}}\big(d+\log\frac{n}{\xi}\big)\sqrt{d+\log\frac{1}{\xi}}\Big) = o(1), \tag{68}$$

~现在我们研究 (66) 中的项 $nT(B^*)^2 I$。注意到，在 E 上且在我们的假设下，$B^* = \frac{2\sqrt{2}B}{2^{r^*}C\mu_{\mathsf{est}}n_{LB}} = O(\frac{K_2}{\mu_{\mathsf{est}}n})$ 由引理 B.2 的性质 3 成立。因此，采用与~上文类似的论证，同样在 E 与 (5) 上，

$$\|nT(B^*)^2 I\| = nT(B^*)^2 = O\Big(nT\cdot\frac{K_2^2}{\mu_{\mathsf{est}}^2 n^2}\Big) = O\Big(\sigma_\varepsilon^2 \max_{i\in[n]}\|\Gamma_i^{-\frac{1}{2}}\|^2 \frac{1}{n\mu_{\mathsf{est}}}\big(d+\log\frac{n}{\xi}\big)\Big) = o(1). \tag{69}$$

^ ^ 因此，为了研究 $V(\beta_{\mathsf{DP}})_{\mathsf{DP}}$ 的收敛性，利用 (66) 以及 (66) 右边第二项和第三项为 $o(1)$ 这一事实，我们最终聚焦于 (66) 右边的第一项：

$$\frac{nT}{n_{\tilde{S}}^2}\sum_{i\in S}(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})^{\mathsf{T}} = \frac{n}{n_{\tilde{S}}}\Big(\frac{T}{n_{\tilde{S}}}\sum_{i\in S}(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})^{\mathsf{T}}\Big). \tag{70}$$

注意，根据 $n_{\tilde{S}} = \max\{|S|,\ n_{LB}\}$ 的定义，

$$1 \leq \frac{n}{n_{\tilde{S}}} \leq \frac{n}{n_{LB}}, \tag{71}$$

并根据 n_{LB} 算法 1 中的定义以及引理 B.2 的性质 1，

$$\frac{n}{n_{LB}} = 1+\frac{n-n_{LB}}{n_{LB}} = 1+\frac{\frac{4}{\mu_{\mathsf{est}}}\sqrt{2R\log\frac{4R}{\xi}}}{n_{LB}} = 1+O\Big(\frac{1}{n\mu_{\mathsf{est}}}\sqrt{R\log\frac{R}{\xi}}\Big). \tag{72}$$

考虑式 (72)，由式 (5) 可知，$\frac{n}{n_{LB}} \to 1$ ~~as~~ $n\to\infty$；根据式 (71)，我们也有 $\frac{n}{n_{\tilde{S}}} \to 1$。然后，由式 (70) 和连续映射定理，只需找到以下量的概率极限

$$\frac{T}{n_{\tilde{S}}}\sum_{i\in S}(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})^{\mathsf{T}}.$$

首先回顾算法 2 第 2 行，$n_{\tilde{S}} = \max(|S|, n_{LB})$ 且 $S = |\{i : \|\hat{}\beta_i - \hat{}m_{r^*-1}\| \leq \frac{B}{2^{r^*}}\}|$。注意，$n_{\tilde{S}} = |S|$ 在 ~E 上：事实上，根据引理 B.2 的性质 3，$|\{i : \|\hat{}\beta_i - \hat{}m_{r^*-1}\| \leq \frac{B}{2^{r^*}}\}| \geq n_{LB}$。因此，我们研究以下表达式的极限

$$\frac{T}{n_{\tilde{S}}}\sum_{i\in S}(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})^{\mathsf{T}} = \frac{T}{|S|}\sum_{i\in S}(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i-\hat{\beta}_{\mathsf{DP}})^{\mathsf{T}}. \tag{73}$$

~ ^–回顾引理 B.2 的性质 3，可知在 E 上有 $\beta_{\mathsf{DP}} = \beta_S + B^* Z$，其中 $Z \sim N($ 0, $I)$ 的贡献有界，$|S| \geq n_{LB} \geq n/2$，且 $B^* = O(\frac{K_2}{n\mu_{\mathsf{est}}})$。据此，我们将 (73) 写为

(74)
$$\frac{T}{|S|}\sum_{i\in S}(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{\mathsf{DP}})^{\mathsf{T}} = \frac{T}{|S|}\sum_{i\in S}(\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)^{\mathsf{T}} - T(\bar{\beta}_S - \beta)(\bar{\beta}_S - \beta)^{\mathsf{T}} + T(B^*)^2 ZZ^{\mathsf{T}}.$$

下文将说明，(74) 右侧的最后两项为 $o(1)$。根据引理 B.2 的性质 4 和 (47)，在~我们的条件下，在 E 上有

(75)
$$\begin{aligned}\|T(\bar{\beta}_S - \beta)(\bar{\beta}_S - \beta)^{\mathsf{T}}\| = T\|\bar{\beta}_S - \beta\|^2 &= TO\Big(K_1^2 + \frac{K_2^2 R\log\frac{R}{\xi}}{n^2\mu_{\mathsf{est}}^2}\Big)\\ &= O\Big(\frac{d+\log\frac{1}{\xi}}{n} + \frac{(d+\log\frac{n}{\xi})R\log\frac{R}{\xi}}{n^2\mu_{\mathsf{est}}^2}\Big) = o(1).\end{aligned}$$

~此外，根据我们先前围绕 (69) 得出的结论，在 E 上有，

(76)
$$\|T(B^*)^2 ZZ^{\mathsf{T}}\| = T(B^*)^2\|Z\|^2 = o(1).$$

因此，根据 (74)，要完成收敛分析，只需证明

(77)
$$\frac{T}{|S|}\sum_{i\in S}(\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)^{\mathsf{T}} \xrightarrow{p} \Big(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}\Big).$$

记作 $A_S := \frac{1}{|S|}\sum_{\overline{i\in}S}($ ^$\beta_i - \beta)($^$\beta_i - \beta)^{\mathsf{T}}$，首先注意到

(78)
$$\frac{T}{|S|}\sum_{i\in S}(\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)^{\mathsf{T}} = \frac{T}{n}\sum_{i=1}^{n}(\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)^{\mathsf{T}} + \frac{T(n-|S|)}{n}(A_S - A_{S^{\mathsf{c}}}).$$

下文将论证 $T_{(n-|S|)\,n}(A_S - A_{S^{\mathsf{c}}}) \to 0$。我们有

$$T\Big\|\frac{n-|S|}{n}(A_S - A_{S^{\mathsf{c}}})\Big\| = \frac{T|S^{\mathsf{c}}|}{n}\|A_S - A_{S^{\mathsf{c}}}\| \leq \frac{T|S^{\mathsf{c}}|}{n}(\|A_s\| + \|A_{S^{\mathsf{c}}}\|) \leq \frac{2T|S^{\mathsf{c}}|}{n}\max_{i\in[n]}\|\hat{\beta}_i - \beta\|^2,$$

~并且根据引理 B.2 的性质 1 和 3，在 E 上，

$$\frac{|S^{\mathsf{c}}|}{n} = \frac{n-|S|}{n} \leq \frac{n-n_{LB}}{n} = O\Big(\frac{1}{n\mu_{\mathsf{est}}}\sqrt{R\log\frac{R}{\xi}}\Big),$$

~而根据引理 4.3，对于足够大的 T，在我们的条件和假设下，在 E 上，

$$\max_{i\in[n]}\|\hat{\beta}_i - \beta\| = O\Big(\sqrt{\frac{1}{T}\big(d+\log\frac{n}{\xi}\big)}\Big).$$

因此，在此事件上，

$$T\Big\|\frac{n-|S|}{n}(A_S - A_{S^{\mathsf{c}}})\Big\| \leq \frac{2T|S^{\mathsf{c}}|}{n}\max_{i\in[n]}\|\hat{\beta}_i - \beta\|^2 = O\Big(\frac{d+\log\frac{n}{\xi}}{n}\sqrt{R\log\frac{R}{\xi}}\Big).$$

因此，根据我们的增长条件 (5)，

$$T\left(\frac{n-|S|}{n}(A_S - A_{S^c})\right) \xrightarrow{p} 0. \tag{79}$$

现在，结合（77）、（78）和（79），可知我们只需考虑 whether$_n^T \sum_{i=1}^n (\hat{}\beta_i - \beta)(\hat{}\beta_i - \beta)^\top \xrightarrow{p} (\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})$。首先，记 $W_{i,T} := \sqrt{T}(\hat{}\beta_i - \beta) = \Gamma_{i,T}^{-1} S_{i,T}$，其中 $\Gamma_{i,T} := \frac{1}{T}\sum_{t=1}^T x_{i,t}x_{i,t}^\top$，以及 $S_{i,T} := \frac{1}{\sqrt{T}}\sum_{t=1}^T x_{i,t}\varepsilon_{i,t}$。因此，剩下需要证明的是以下内容：

$$\frac{T}{n}\sum_{i=1}^n W_{i,T}W_{i,T}^\top \xrightarrow{p} \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}\right). \tag{80}$$

根据斯卢茨基定理和标准论证（所需的要素例如可参见定理 4.6 的证明），我们知道 $W_{i,T} \xrightarrow{d} N(0, \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})$，当 $T\to\infty$ 时成立；由连续映射定理可推出 $W_{i,T}W_{i,T}^\top \xrightarrow{d} Z_iZ_i^\top$，其中 $Z_i \sim N(0, \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})$。根据假设 4.4，我们可以据此得出结论：当 $T\to\infty$ 时，$\mathbb{E}[W_{i,T}W_{i,T}^\top] \to \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}$。事实上，对于某个 $\gamma \in (0, \frac{\delta}{2})$，其中 δ 如假设 4.4 中所述，令 $p = {}^{1+\delta}/_{1+\gamma}{}^2 > 1$ 和 $q = \frac{p}{p-1}$，根据 Hölder 不等式，

$$\sup_{T\ge T_0}\mathbb{E}\|W_{i,T}W_{i,T}^\top\|^{1+\gamma} \le \left(\sup_{T\ge T_0}\mathbb{E}\|\Gamma_{i,T}^{-1}\|^{2(1+\gamma)p}\right)^{1/p}\left(\sup_{T\ge T_0}\mathbb{E}\|S_{i,T}\|^{2(1+\gamma)q}\right)^{1/q}.$$

请注意，上式右侧的第一项由假设 4.4 界定，第二项由引理A.15界定。因此，我们已证明在$W_{i,T}W_{i,T}^\top$上一致可积。现在可以证明最后的结果；也就是说，我们要证明（80）。为此，请注意

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n W_{i,T}W_{i,T}^\top - \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}\right)\\
&\quad= \frac{1}{n}\sum_{i=1}^n \left(W_{i,T}W_{i,T}^\top - \mathbb{E}[W_{i,T}W_{i,T}^\top]\right) + \frac{1}{n}\sum_{i=1}^n \left(\mathbb{E}[W_{i,T}W_{i,T}^\top] - \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}\right)\\
&\quad\quad + \frac{1}{n}\sum_{i=1}^n \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1} - \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}\right).
\end{aligned} \tag{81}$$

将上式右侧的各项分别记为 I_1、I_2 和 I_3，我们将证明每一项都依概率收敛于 0，从而完成证明。

首先注意到，当 $\delta < 1$ 足够小时，使得假设 4.4 在 2δ下成立时，I_1 式（81）在 $L^{1+\delta}$ 中收敛于 0。因此，它也依概率收敛于 0，因为

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n (W_{i,T}W_{i,T}^\top - \mathbb{E}[W_{i,T}W_{i,T}^\top])\right\|^{1+\delta} &\le \mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n (W_{i,T}W_{i,T}^\top - \mathbb{E}[W_{i,T}W_{i,T}^\top])\right\|_F^{1+\delta}\\
&\le \sum_{j=1}^d\sum_{k=1}^d \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n \Psi_{i,T,j,k}\right|^{1+\delta},
\end{aligned}$$

其中，我们将 $\Psi_{i,T,j,k}$ 定义为中心化的 $W_{i,T}W_{i}^{\mathsf{T}},_{T}$ 的第（j,k）个元素。现在我们证明上述右侧为 $o(1)$。

$$\begin{aligned}\sum_{j=1}^{d}\sum_{k=1}^{d}\mathbb{E}\Big|\frac{1}{n}\sum_{i=1}^{n}\Psi_{i,T,j,k}\Big|^{1+\delta} &\overset{(1)}{\leq} \frac{C}{n^{1+\delta}}\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{k=1}^{d}\mathbb{E}|\Psi_{i,T,j,k}|^{1+\delta}\\ &\leq \frac{C'd^2}{n^{\delta}}\sup_{T\geq T_0}\sup_{i,j,k}\mathbb{E}|\Psi_{i,t,j,k}|^{1+\delta}\\ &\leq \frac{C''d^2}{n^{\delta}}\sup_{i,T\geq T_0}\mathbb{E}\Big\|W_{i,T}W_{i,T}^{\mathsf{T}}-\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]\Big\|^{1+\delta}\overset{(2)}{=}o(1),\end{aligned}$$

在上述内容中，C，C'，$C''>0$是与 n和 T无关的常数，不等式 (1) 由引理 A.9 得出，而等式 (2) 由假设 4.1、4.2 和 4.4 得出。

接下来，考虑 (81) 中的项 I_2，

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]-\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1})\Big\|\leq\sup_{i\geq1}\Big\|\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]-\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}\Big\|.$$

因此，我们将证明当 $T\to\infty$时，$\sup_{i\geq1}\|\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]-\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}\|\to0$。我们可以写成

$$\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]-\Gamma_i^{-1}\Lambda_i\Gamma_i^{-1}=\big(\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]-\Gamma_i^{-1}\Lambda_{i,T}\Gamma_i^{-1}\big)+\Gamma_i^{-1}(\Lambda_{i,T}-\Lambda_i)\Gamma_i^{-1},\tag{82}$$

对于 $\Lambda_{i,\ T}:=\mathbb{E}[S_{i,T}S_{i,T}^{\mathsf{T}}]$。考虑 (82) 右侧的第二项。回想，

$$\Lambda_{i,T}-\Lambda_i=\sum_{h=-(T-1)}^{T-1}\Big(1-\frac{|h|}{T}\Big)\mathbb{E}[x_{i,0}x_{i,h}^{\mathsf{T}}\varepsilon_{i,0}\varepsilon_{i,h}]-\sum_{h=-\infty}^{\infty}\mathbb{E}[x_{i,0}x_{i,h}^{\mathsf{T}}\varepsilon_{i,0}\varepsilon_{i,h}],$$

并且，记 $\Gamma_i(h)$ 为 $:=\mathbb{E}[x_{i,0}x_{i,h}^{\mathsf{T}}\varepsilon_{i,0}\varepsilon_{i,h}]$，我们有

$$\begin{aligned}\sup_{i\geq1}\|\Gamma_i^{-1}(\Lambda_{i,T}-\Lambda_i)\Gamma_i^{-1}\|&\leq\frac{1}{c_\Gamma^2}\sup_{i\geq1}\|\Lambda_{i,T}-\Lambda_i\|\\ &\leq\frac{1}{c_\Gamma^2}\Big(\sum_{|h|\geq T}\sup_{i\geq1}\|\Gamma_i(h)\|+\frac{1}{T}\sum_{|h|<T}|h|\sup_{i\geq1}\|\Gamma_i(h)\|\Big)\to0,\end{aligned}$$

因为次高斯性和混合常数在 $i\in\mathbb{N}$ 中均一致。特别地，由引理 A.11 和假设 4.1-4.2，存在常数 $C>0$，使得 $\sup_{i\geq1}\|\Gamma_i(h)\|\leq C|h|^{-3-q/2}$，因此上述求和为 $O(T^{-1})$。考虑 (82) 右侧的第一项。我们将证明 $\sup_{i\geq1}\|\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]-\Gamma_i^{-1}\Lambda_{i,\ T}\Gamma_i^{-1}\|\to0$ 当 $T\to\infty$时成立。令 $\Delta_{i,\ T}:=\Gamma_{i}^{-1},_{T}-\Gamma_i^{-1}$。则

$$\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]-\Gamma_i^{-1}\Lambda_{i,T}\Gamma_i^{-1}=\mathbb{E}[\Delta_{i,T}S_{i,T}S_{i,T}^{\mathsf{T}}]\Gamma_i^{-1}+\Gamma_i^{-1}\mathbb{E}[S_{i,T}S_{i,T}^{\mathsf{T}}\Delta_{i,T}]+\mathbb{E}[\Delta_{i,T}S_{i,T}S_{i,T}^{\mathsf{T}}\Delta_{i,T}],$$

从而

$$\begin{aligned}&\sup_{i\geq1}\|\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf{T}}]-\Gamma_i^{-1}\Lambda_{i,T}\Gamma_i^{-1}\|\\ &\quad\leq2\frac{1}{c_\Gamma}\sup_{i\geq1}(\mathbb{E}\|S_{i,T}\|^{2q})^{1/q}\sup_{i\geq1}(\mathbb{E}\|\Delta_{i,T}\|^{p})^{1/p}+\sup_{i\geq1}(\mathbb{E}\|S_{i,T}\|^{2q'})^{1/q'}\sup_{i\geq1}(\mathbb{E}\|\Delta_{i,T}\|^{2p'})^{1/p'}\end{aligned}$$

对于 p，q，p'，q'，使得 $\frac{1}{p}+\frac{1}{q}=1$ 和 $\frac{1}{p'}+\frac{1}{q'}=1$。现在，对于任意 $r>1$，我们有

$$\mathbb{E}\|\Delta_{i,T}\|^r=\mathbb{E}\|\Gamma_i^{-1}{}_{,T}-\Gamma_i^{-1}\|^r\le\|\Gamma_i^{-1}\|^r\big(\mathbb{E}\|\Gamma_i^{-1}{}_{,T}\|^{ar}\big)^{1/a}\big(\mathbb{E}\|\Gamma_{i,T}-\Gamma_i\|^{br}\big)^{1/b},$$

满足 $\frac{1}{a}+\frac{1}{b}=1$。现在，取 $p=q=2$、$p'=1+\delta/4$、$q'=1+4/\delta$、$a=\frac{2+\delta}{2+\delta}$ 以及 $b=2+4/\delta$，其中 $\delta>0$ 如假设 4.4 所述，由引理A.15可得 $\sup_{i\ge1}(\mathbb{E}\|S_{i,T}\|^4)^{1/2}=O(1)$ 以及 $\sup_{i\ge1}(\mathbb{E}\|S_{i,T}\|^{2+8/\delta})^{1/(4+\delta)}=O(1)$。类似地，对于我们对 a、b、p、p' 的取值，有 $ap=2(2+\delta)/(2+\delta)$、$bp=4+\frac{8}{\delta}$、$2ap'=2+\delta$ 以及 $2bp'=2b(1+\delta/4)$。因此，根据假设 4.4，有 $\sup_{i\ge1}\mathbb{E}\|\Gamma_i^{-1}{}_{,T}\|^{ap}=O(1)$ 以及 $\sup_{i\ge1}\mathbb{E}\|\Gamma_i^{-1}{}_{,T}\|^{2ap'}=O(1)$。此外，由引理A.15可得 $\sup_{i\ge1}\mathbb{E}\|\Gamma_{i,T}-\Gamma_i\|^{bp}=O(T^{-2-\frac{4}{\delta}})$ 以及 $\sup_{i\ge1}\mathbb{E}\|\Gamma_{i,T}-\Gamma_i\|^{2bp'}=O(T^{-b(1+\delta/4)})$，从而可以得出 $\sup_{i\ge1}\|\mathbb{E}[W_{i,T}W_{i,T}^{\mathsf T}]-\Gamma_i^{-1}\Lambda_i{}_T\Gamma_i^{-1}\|=o(1)$，进而 $I_2\to0$ as $(n,T)\to\infty$。

最后，I_3 根据假设，当 $n\to\infty$ 时趋于零，且与 T 独立。□

Appendix E. Private Covariance Estimation with Two Groups

定理 E.1. 令 $\hat\beta_{1,\mathsf{DP}}$ 按照定理 *4.8* 中的方法，使用 μ_{est}-*uGDP* 进行估计。设 $(\beta_i,z_i)_{i=1}^n$ 和 $\hat\beta_{1,\mathsf{DP}}$ 是算法 *5* 的输入，其隐私级别为 μ_{var}，并将所得估计记为 $\hat V(\hat\beta_{1,\mathsf{DP}})_{\mathsf{DP}}$。

(a) 隐私。$(\hat\beta_{1,\mathsf{DP}},\hat V(\hat\beta_{1,\mathsf{DP}})_{\mathsf{DP}})$ 满足 $\sqrt{\mu_{\mathsf{est}}^2+\mu_{\mathsf{var}}^2}$-*uGDP*。***(b)*** 收敛。若满足定理 *4.8* 的条件，则当 $(n,T)\to\infty$，

$$\big(\hat V(\hat\beta_{1,\mathsf{DP}})_{\mathsf{DP}}\big)^{-\frac12}(\hat\beta_{1,\mathsf{DP}}-\beta)\xrightarrow{d}N(0,I_d).$$

证明。对于证明中的隐私部分，只需验证统计量 $\frac{1}{n_{\tilde S}^2}\sum_{i\in S}(\hat\beta_i-\hat\beta_{1,\mathsf{DP}})(\hat\beta_i-\hat\beta_{1,\mathsf{DP}})^{\mathsf T}$ 的敏感度。注意，根据后处理（命题 2.6），在进行敏感度分析时，可将 $\beta_{1,\mathsf{DP}}$ 视为固定量。考虑两个相邻数据集 D 和 D′，不失一般性地假设两个数据集之间仅第一位用户的数据不同。因此，数据集 D 和 D′ 之间的唯一差异在于 (β_1,z_1) 与 (β_1',z_1') 不同。若 $z_1=z_1'$，则分析可归结为对 (63) 的分析，敏感度至多为 $4\kappa^2/n_{LB}^2$。若 $z_1=z_1'=0$，则敏感度为 0。因此，唯一需要考虑的新情形是 $z_1=1$ 且 $z_1=0$（相反的情形可由对称性得出）。此时 $1\notin S'$，且 $1\in S$ 或 $1\notin S'$。前一种情形下敏感度为 0，而后一种情形下，根据 (63)，敏感度不超过 $4\kappa^2/n_{LB}^2$。因此，该统计量的敏感度全局有界于 $4\kappa^2/n_{LB}^2$，进而根据我们在 (64) 附近的讨论以及组合（命题 2.7），可得 that $(\beta_{\mathsf{DP}},V(\beta_{1,\mathsf{DP}})_{\mathsf{DP}})$ 满足 μ_{var}-uGDP。由于 $V(\beta_{1,\mathsf{DP}})_{\mathsf{DP}}$ 是 $\hat V(\beta_{1,\mathsf{DP}})_{\mathsf{DP}}$ 的确定性后处理，根据命题 2.6，隐私结果即得证。

效用证明与定理 5.1 的效用证明几乎完全相同。与 (68) 中一样，在我们的高概率事件上有 $\|nTW\|=o(1)$，因为同样有 $\kappa=O(K_2)$，并且正如我们在 (62) 中所示，我们再次有 $n_{LB}=\Omega(\sum_{i=1}^n z_i)=\Omega(n)$。出于同样的原因，与 (69) 中一样，在我们的高概率事件上还有 $\|nT(B^*)I\|=o(1)$。因此，我们只需分析

$$\frac{Tn}{n_{\tilde S}^2}\sum_{i\in S}(\hat\beta_i-\hat\beta_{1,\mathsf{DP}})(\hat\beta_i-\hat\beta_{1,\mathsf{DP}})^{\mathsf T},$$

并且与 (72) 中一样，在高概率事件 (61) 上有 $\frac{n}{S} \to 1$，其中 $n \to \infty$。此外，与引理 B.2 的性质 3 中一样，我们有 $|S| \geq n_{LB}$，因此 $n_{\tilde{S}} = |S|$。由此归结为推导以下量的概率极限

$$\frac{T}{|S|} \sum_{i \in S} (\hat{\beta}_i - \hat{\beta}_{1,\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{1,\mathsf{DP}})^\mathsf{T},$$

并且由于 (74)、(75)、(76)、(77)、(78) 和 (79) 在我们的情形下均适用，我们可以等价地得到..... 的概率极限

$$\frac{T}{\sum_{i=1}^n z_i} \sum_{i:z_i=1} (\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)^\mathsf{T},$$

并且，只需在我们的设定中考虑样本 $\{i \in [n] : z_i = 1\}$ 而非 $\{i \in [n]\}$，(80) 后的讨论便可得出，

$$\frac{T}{\sum_{i=1}^n z_i} \sum_{i:z_i=1} (\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)^\mathsf{T} \xrightarrow{p} \lim_{n\to\infty} (\frac{1}{pn} \sum_{i:z_i=1} \Gamma_i^{-1} \Lambda_i \Gamma_i^{-1}),$$

从而证明了我们所需的结果。 □

自然，定理 E.1 也为算法在输入 $(\beta_i,\ z_i)_{i=1}^n$ 时的输出提供保证，从而为 $V(\beta_{0,\ \mathsf{DP}})_{\mathsf{DP}}$ 提供保证。因此，我们可以得到以下推论。由此，也可以像推论 5.2 和 5.3 那样，为系数向量 θ 得到私有且渐近有效的置信区间和 Wald 检验。

^ ^ ^ ^ ^ ^ 推论 *E.2.* 设 $(\beta_{1,\mathsf{DP}}, V(\beta_{1,\mathsf{DP}})_{\mathsf{DP}})$ 和 $(\beta_{0,\mathsf{DP}}, V(\beta_{0,\mathsf{DP}})_{\mathsf{DP}})$ 如定理 *E.1* 所述。则 $(\hat{}\beta_1, \hat{}\beta_0, \hat{}V(\hat{}\beta_{1,\mathsf{DP}})_{\mathsf{DP}}, \hat{}V(\hat{}\beta_{0,\mathsf{DP}})_{\mathsf{DP}})$ 满足 $\sqrt{2(\mu_{\mathsf{est}}^2 + \mu_{\mathsf{var}}^2)}$-*uGDP*。若进一步满足定理 *E.1* 的条件，则，

$$\left(\hat{V}(\hat{\beta}_{1,\mathsf{DP}})_{\mathsf{DP}} + \hat{V}(\hat{\beta}_{0,\mathsf{DP}})_{\mathsf{DP}}\right)^{-\frac{1}{2}} \left(\hat{\beta}_{1,\mathsf{DP}} - \hat{\beta}_{0,\mathsf{DP}} - \gamma\right) \xrightarrow{d} N(0, I_d).$$

算法 5 处理组用户级协方差矩阵估计

^1:**输入：** 用户级 OLS 估计 β_i、组成员变量 z_i、算法 4 的输出 $\hat{}(\beta_{1、\ \mathsf{DP}}、\ \hat{}m_{r^*-1}、r^*、n_{LB}、B^*)$，以及协方差隐私参数 μ_{var}。2: 令 $\delta_{\mathsf{DP}} \leftarrow \|\hat{}\beta_{1、\ \mathsf{DP}} - \hat{}m_{r^*-1}\|$，$\kappa \leftarrow \frac{B}{2^{r^*}} + \delta_{\mathsf{DP}}$，$S \leftarrow \{i \in [n] : \|\hat{}\beta_i - \hat{}m_{r^*-1}\| \leq \frac{B}{2^{r^*}} z_i\}$，$n_{\tilde{S}} \leftarrow \max(|S|、n_{LB})$ 3: 令

$$\tilde{V}(\hat{\beta}_{1,\mathsf{DP}})_{\mathsf{DP}} \leftarrow \frac{1}{n_{\tilde{S}}^2} \sum_{i \in S} (\hat{\beta}_i - \hat{\beta}_{1,\mathsf{DP}})(\hat{\beta}_i - \hat{\beta}_{1,\mathsf{DP}})^\mathsf{T} + (B^*)^2 I + W,$$

其中 $W_{i,\ i} \overset{iid}{\sim} N(0, \frac{16\kappa^4}{n_{LB}^4 \mu_{\mathsf{var}}^2})$ 和 $W_{i,\ j} \overset{iid}{\sim} N(0, \frac{8\kappa^4}{n_{LB}^4 \mu_{\mathsf{var}}^2})$ 对于 $j > i$ 且有 $W_{i,\ j} = W_{j,\ i}$ 对于 $j < i$。4: 令 $\hat{}V(\hat{}\beta_{1,\ \mathsf{DP}})_{\mathsf{DP}}$ 表示 $\tilde{V}(\hat{}\beta_{1,\ \mathsf{DP}})_{\mathsf{DP}}$ 在锥 $\mathbf{M}_{symm}^{d\times d}$ 上的投影。 ^ ^5: **输出空间：** $V(\beta_{1,\ \mathsf{DP}})_{\mathsf{DP}}$

Appendix F. Examples

F.1. 示例 1。

证明。 在这种情况下

$$\Gamma = \begin{bmatrix} 1 & m^\top \\ m & mm^\top \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{0}^\top \\ \mathbf{0} & \sum_{j=0}^{\infty} \Psi_j \Sigma_e \Psi_j^\top \end{bmatrix}.$$

此外，由于 $e_{i,\,t} \overset{iid}{\sim} \mathrm{subG}(\sigma_x^2)$，我们有 $\Gamma^{-\frac{1}{2}} x_{i,\,t} \sim \mathrm{subG}(\sigma_x^2 \|\Gamma^{-\frac{1}{2}}\|^2 \sum_{j=0}^{\infty} \|\Psi_j\|^2)$。其余条件很容易验证，但请注意

$$\lambda_{\min}(\Gamma) \geq \frac{\lambda_{\min}(\sum_{j\geq 0} \Psi_j \Sigma_e \Psi_j^\top)}{1 + \|m\|^2 + \lambda_{\min}(\sum_{j\geq 0} \Psi_j \Sigma_e \Psi_j^\top)}, \quad \text{and} \quad \lambda_{\max}(\Gamma) \leq 1 + \|m\|^2 + \max\{1, \Sigma_e \sum_{j\geq 0} \|\Psi_j\|^2\},$$

这些条件均由我们的示例假设所界定。指数强混合性由（Mokkadem，1988年，定理1）给出。□

F.2. 例2。

证明。 要看出这一点，请注意因果性假设意味着 $\sum_{j=0}^{\infty} |\psi_j| < \infty$，（例如 Brockwell 和 Davis 1991年），因此，对于任意 $\lambda \in \mathbb{R}$,

$$\mathbb{E}[\exp(\lambda \varepsilon_{i,t})] = \prod_{j=0}^{\infty} \mathbb{E}[\exp(\lambda \psi_j r_{i,t-j})] \leq \prod_{j\geq 0} \exp\left(\frac{\sigma_r^2 \lambda^2 \psi_j^2}{2}\right) = \exp\left(\frac{\sigma_r^2 \lambda^2}{2} \sum_{j\geq 0} \psi_j^2\right),$$

因此，其代理方差至多为 $\sigma^2 \|\boldsymbol{\psi}\|_{\ell_2}^2$。此外，Mokkadem (1988)证明，具有绝对连续创新项的因果 ARMA 模型是指数强混合的。最后，$(\varepsilon_{i,t})_{t\geq 1}$ 的谱密度具有以下形式

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \sum_{k=1}^{p} \phi_k e^{-ik\lambda}|^2}, \quad \lambda \in [-\pi, \pi],$$

其中 $\sigma^2 := \mathrm{Var}(r_{i,t})$ (Brockwell 和 Davis, 1991年)，因果性意味着分母的所有根都位于单位圆之外，因此，记作 $\Phi(z) = 1 - \sum_{k=1}^{p} \phi_k z^p$,

$$0 < \frac{\sigma^2}{2\pi \max_\lambda |\Phi(e^{-i\lambda})|^2} \leq f(\lambda) \leq \frac{\sigma^2}{2\pi \min_\lambda |\Phi(e^{-i\lambda})|^2} < \infty.$$

□

F.3. 示例 3。

证明。 对于任意 $v \in \mathbb{S}^{d-1}$，都有 $(v^\top x_{1,t}, \ldots, v^\top x_{1,T})$ 服从高斯分布。此外，协方差矩阵 $\Sigma_T(v)$ 具有如下形式（参见 Brockwell 和 Davis 1991年）

$$\lambda_{\min}(\Sigma_T(v)) \geq \frac{\lambda_{\min}(\Sigma_z)}{\sup_\omega s_{\max}^2(P(e^{-i\omega}))},$$

其中 $P(e^{-i\omega}) := I - \sum_{k=1}^{q} A_k e^{-i\omega k}$ 根据因果性假设具有有界谱，且 $s_{\max}(P(e^{-i\omega}))$ 是 $P(e^{-i\omega})$ 的最大奇异值。因此，由于 $\lambda_{\min}(\Sigma_z) > 0$

（非退化创新情形），存在某个 $c>0$，使得 $\lambda_{\min}(\Sigma_z)$ 上确界$_\omega s^2_{\max}(P(e^{-i\omega}))\geq c$，且与 T 和 v 无关。由此可得

$$f_{v^\top x_{1,t},\ldots,v^\top x_{1,T}}(w)\leq(2\pi c)^{-T/2}=\Big(\frac{1}{\sqrt{2\pi c}}\Big)^T,$$

因此我们可以取 $K:=\sqrt{\frac{1}{2\pi c}}$。□