URBAN INSTITUTE · ELEVATE · THE · DEBATE

Tuesday, February 10th, 2026

Benchmarking DP Linear Regression Methods for Statistical Inference

Aaron R. Williams, Andres Barrientos, Joshua Snoke, Claire McKay Bowen

Funded by the Alfred P. Sloan Foundation and
National Science Foundation National Center for Science and Enginering Statistics

Research Question

When do differentially private linear regression methods provide results for statistical inference that are sufficiently accurate to be useful?

Importance

- Differential privacy is one tool data stewards are exploring to tune the tradeoff between expanding data access and protecting confidentiality.
- It is important to understand when DP linear regression does and doesn't work well **for inference**.
- The Urban Institute is working to add linear regression to its validation server.

Previous Work

- "A Feasibility Study of Differentially Private Summary Statistics and Regression Analyses with Evaluations on Administrative and Survey Data" (Barrientos et al., 2024)
  1. First study to focus on inference for DP regression instead of prediction.
  2. What causes the gap between theoretical expectations and empirical performance of DP regression?

Our Contribution

1. We develop a framework for simulating data and evaluating DP linear regression methods for inference under known population parameters.
2. We evaluate violations of assumptions for multiple linear regression with normal errors.

Takeaways

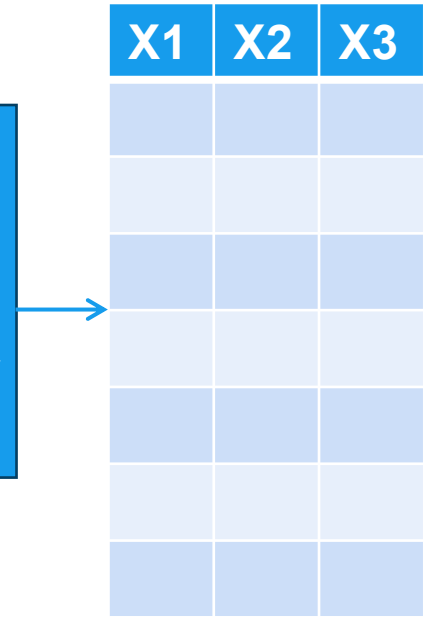
1. Context drives performance of DP regression.
2. Sample size, privacy-loss budget, and sensitivity matter.
3. Mechanisms and techniques for bounds matter but not as much as #2.
4. Violations of MLR assumptions matter but not as much as #2 or #3.
5. Framework is useful pre-testing and evaluation.

Methods

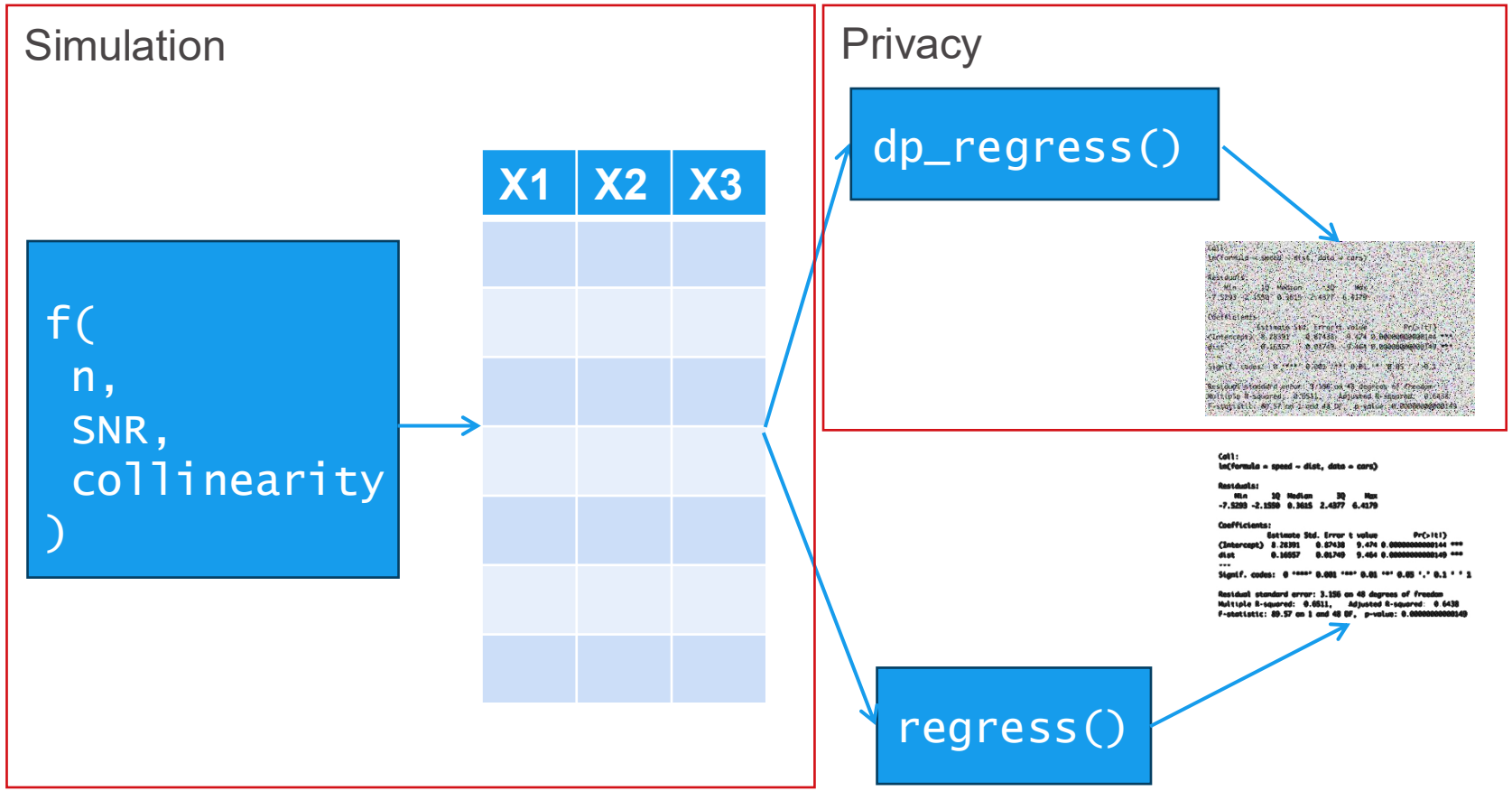
Framework

Simulation

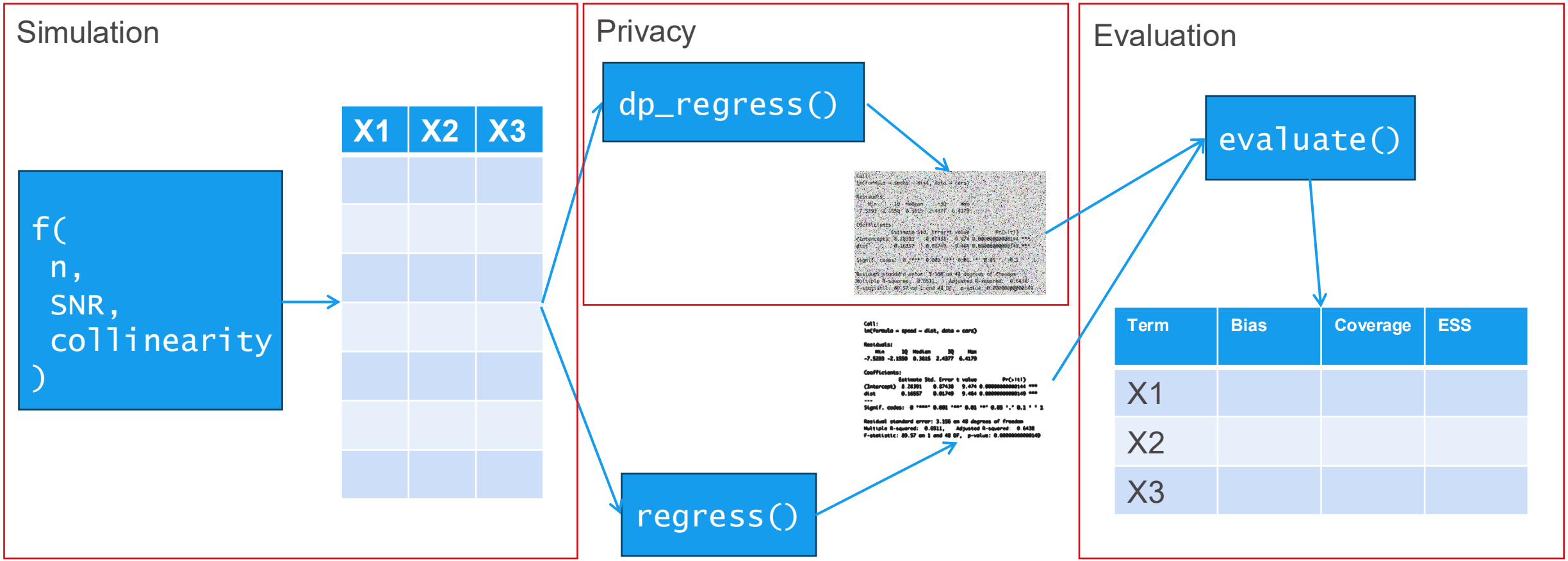
```
f(
 n,
 SNR,
 collinearity
)
```



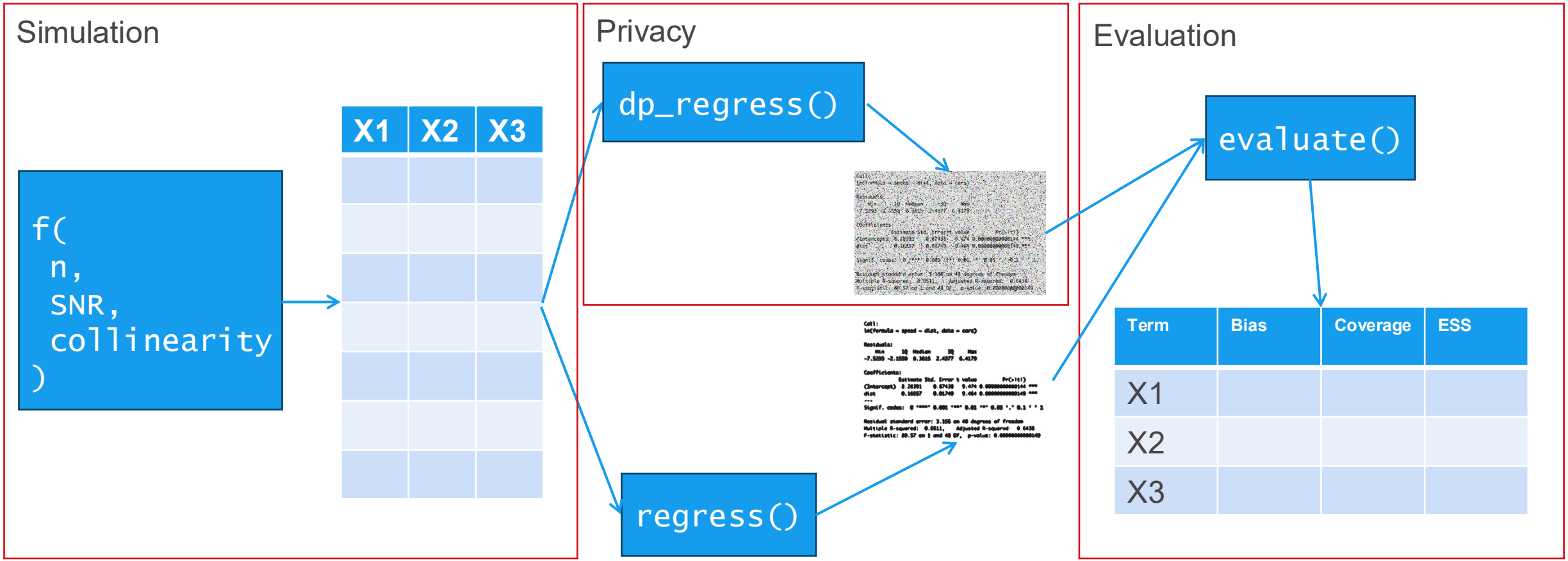
Framework


Simulation
f(
n,
SNR,
collinearity
)
X1
X2
X3
Privacy
dp_regress()
regress()

Framework


Simulation
f(
n,
SNR,
collinearity
)
X1
X2
X3
Privacy
dp_regress()
regress()
Evaluation
evaluate()
Term
Bias
Coverage
ESS
X1
X2
X3

Framework



Specify different scenarios

Repeat many times

Simulation

Data

Y	V1	V2	V3	V4
-0.5	0.2	2.2	c	e
2	0.1	0.3	b	e
0.1	-0.5	0.4	a	d
0.2	1.4	-1	a	e
-0.4	-2.3	-0.1	c	d

Simulation

Data

Y	V1	V2	V3	V4
-0.5	0.2	2.2	c	e
2	0.1	0.3	b	e
0.1	-0.5	0.4	a	d
0.2	1.4	-1	a	e
-0.4	-2.3	-0.1	c	d

$$\vec{Y} = \boldsymbol{X}\vec{\beta} + \vec{e}$$

Simulation

Data

Y	V1	V2	V3	V4
-0.5	0.2	2.2	c	e
2	0.1	0.3	b	e
0.1	-0.5	0.4	a	d
0.2	1.4	-1	a	e
-0.4	-2.3	-0.1	c	d

Y
-0.5
2
0.1
0.2
-0.4

Design Matrix

	X1	X2	X3	X4	X5
1	0.2	2.2	0	1	1
1	0.1	0.3	1	0	1
1	-0.5	0.4	0	0	0
1	1.4	-1	0	0	1
1	-2.3	-0.1	0	1	0

β0
β1
β2
β3
β4
β5

e
0.1
-0.1
0.2
1
0.5

$$\vec{Y} = \boldsymbol{X}\vec{\beta} + \vec{e}$$

Simulation

- N = 100,000
- $rep_{baseline}$ = 100 and $rep_{alternative}$ = 20
- Signal-to-Noise Ratio (SNR) = {0, 0.01, 0.1 0.5, 1}

$$\text{SNR} = \frac{\text{Var}(X_i\beta)}{\sigma^2}.$$

- Vary population coefficients and variances to fix SNR and variable bounds.

Privacy

- We implement the Laplace mechanism and analytic Gaussian mechanism to privatize the sufficient statistics for the regression model. This mechanism adds noise to $\boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{X}^T y$ to ensure DP for a given sensitivity and privacy-loss budget.
- We use a parametric bootstrap to generate confidence intervals that account for DP error and sampling error.

Privacy

- We need to know the sensitivity of the sufficient statistics.
- We place bounds on the continuous variables, $\vec{Y}, \overrightarrow{X_1}, \overrightarrow{X_2}$, to calculate the sensitivity.
    1. Grouped local bounds: use observed minimum and maximum with a specification for each variable.
    2. DP range bounds: estimate noisy (ε = 1) extreme percentiles (0.001, 0.999) to approximate the minimum and maximum for each variable.

Evaluation

Metric	Description	Calculation
Coverage	Proportion of times to confidence interval contains the population parameter	$Coverage_j = \frac{1}{N}\sum_{k}^{N} I(\beta_j \in CI(\widehat{\beta_{j,k}^{DP}}))$
Relative CI Length	Length of the noisy confidence interval relative to the original confidence interval.	$Relative\ CI\ Length_j = \frac{1}{N}\sum_{k=1}^{N} \frac{Length(CI(\widehat{\beta_{j,k}^{DP}}))}{Lengt(CI(\widehat{\beta_{j,k}}))}$
Bias	Relative difference between the average noisy estimate and the estimate without noise	$Relative\ Bias_j = \frac{(\frac{1}{N}\sum_{k=1}^{N}\widehat{\beta_{k,j}^{DP}}) - \widehat{\beta_j}}{\beta_j}$
Power	The probability that a test correctly rejects a null hypothesis when there is an effect.	$Power_j = \frac{1}{N}\sum_{k}^{N} I(0 \notin CI(\beta_{j,k}^{DP}))$
Effective Sample Size	Sample size accounting for the additional variance in the estimator from the noise mechanism	$n_{ESS,j} = n\ \frac{Var(\widehat{\beta_j})}{Var(\widehat{\beta_j^{DP}})}$

Simulation

Alternative Scenario	Parameter	Change from Baseline	Expected Impact without Noise
Imbalanced categorical predictor	Not shown	…	…
Collinearity	σ_{X_1,X_2}	0, 0.5, 0.9	Increased variance
Heteroscedasticity	Not shown	…	…
Skewed Residuals	Not shown	…	…
Sample size	NA	…	…
Omitted variable	NA	…	…

Results

Takeaway #2: Sample size, privacy-loss budget, and sensitivities matter

- *Increasing the sample size improves the coverage, CI lengths, and power.*

Sample Size Drives Results

Analytic Gaussian Mechanism, Bootstrap CIs, Epsilon == 1, DP Range Bounds

sample_size	rep	coverage_probability	relative_length	bias	power
Signal-to-noise ratio: 0.01					
1,000	120	45%	17,322%	−0.0253	1%
10,000	120	44%	7,894%	0.0088	2%
100,000	120	97%	155%	0.0002	100%
Signal-to-noise ratio: 0.1					
1,000	120	45%	349,874%	−2.2945	2%
10,000	120	37%	5,990%	−0.2157	3%
100,000	120	98%	199%	0.0006	100%
Signal-to-noise ratio: 0.5					
1,000	120	22%	123,811%	−1.0834	2%
10,000	120	36%	16,615%	−0.1068	4%
100,000	120	96%	263%	−0.0055	98%

Takeaway #2: Sample size, privacy-loss budget, and sensitivities matter

- *Increasing ε improves the coverage, CI lengths, power, and effective sample size.*

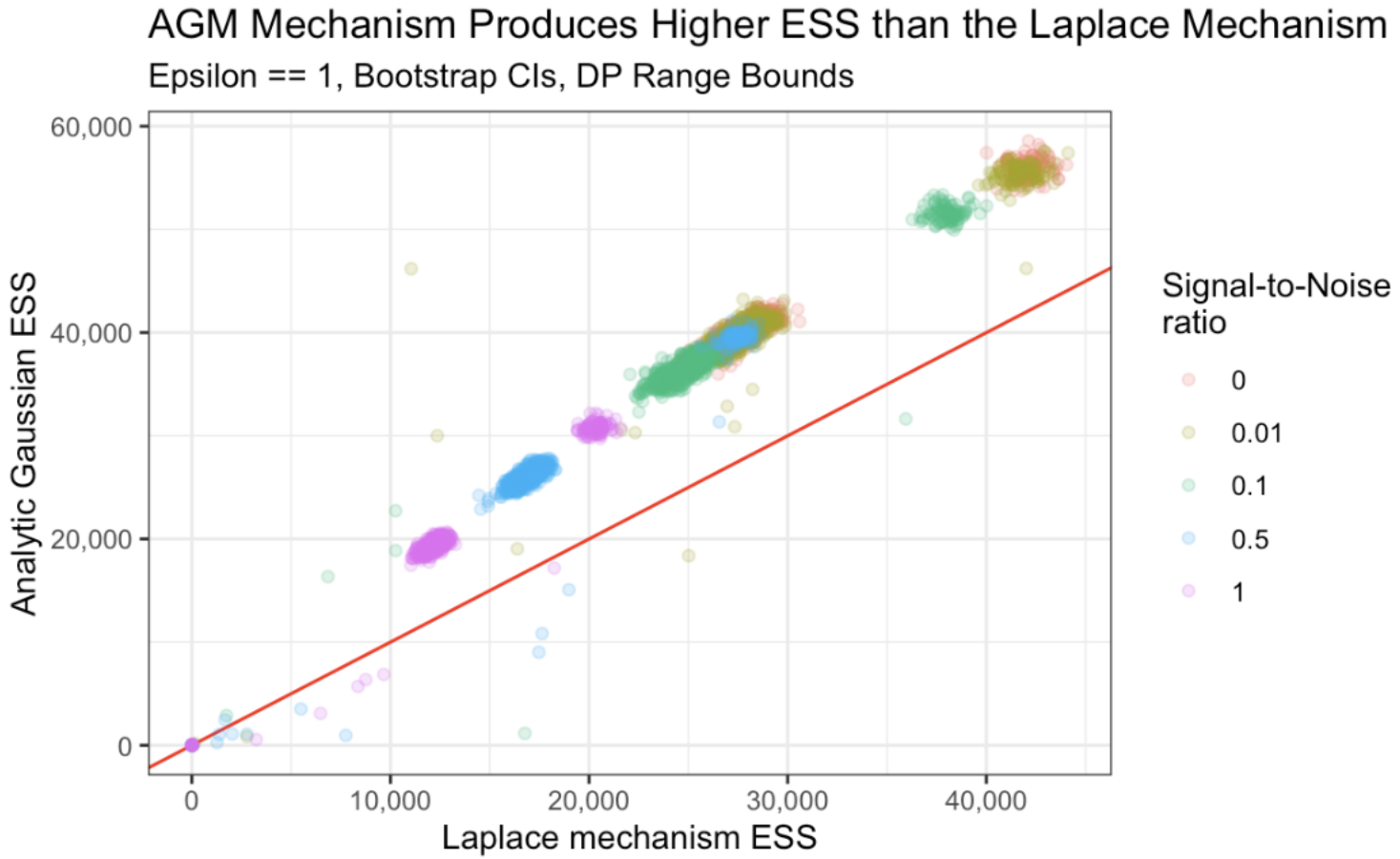
Epsilon Drives Results

Analytic Gaussian Mechanism, Bootstrap CIs, SNR == 0.01

epsilon	rep	coverage_probability	relative_length	bias	power	ess
0.5	600	94%	261%	0.0003	96%	16,483
1	600	96%	161%	0.0007	100%	41,856
5	600	95%	105%	0.0001	100%	93,107
10	600	95%	102%	0.0001	100%	97,743
1,000,000	600	95%	100%	0.0002	100%	100,257

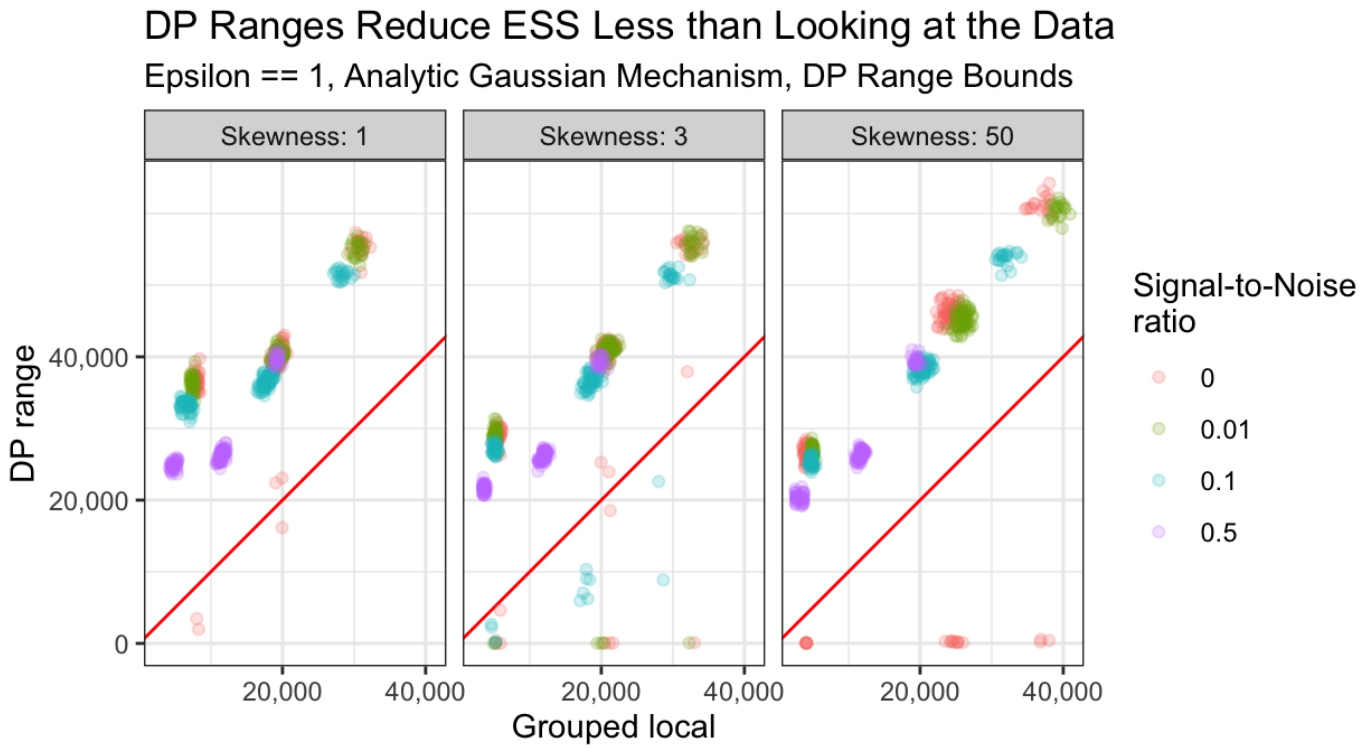
Takeaway #3: Mechanisms and techniques for bounds matter but not as much as #2

- *The Analytic Gaussian Mechanism consistently produces higher effective sample sizes than the Laplace mechanism.*



Takeaway #3: Mechanisms and techniques for bounds matter but not as much as #2

- *DP range bounds consistently produces higher effective sample sizes than the grouped local bounds.*



Takeaway #4: Violations of MLR assumptions affect results but not as much as #2 or #3.

- *Omitting x2_continuous does not introduce extra bias and decreases variance.*

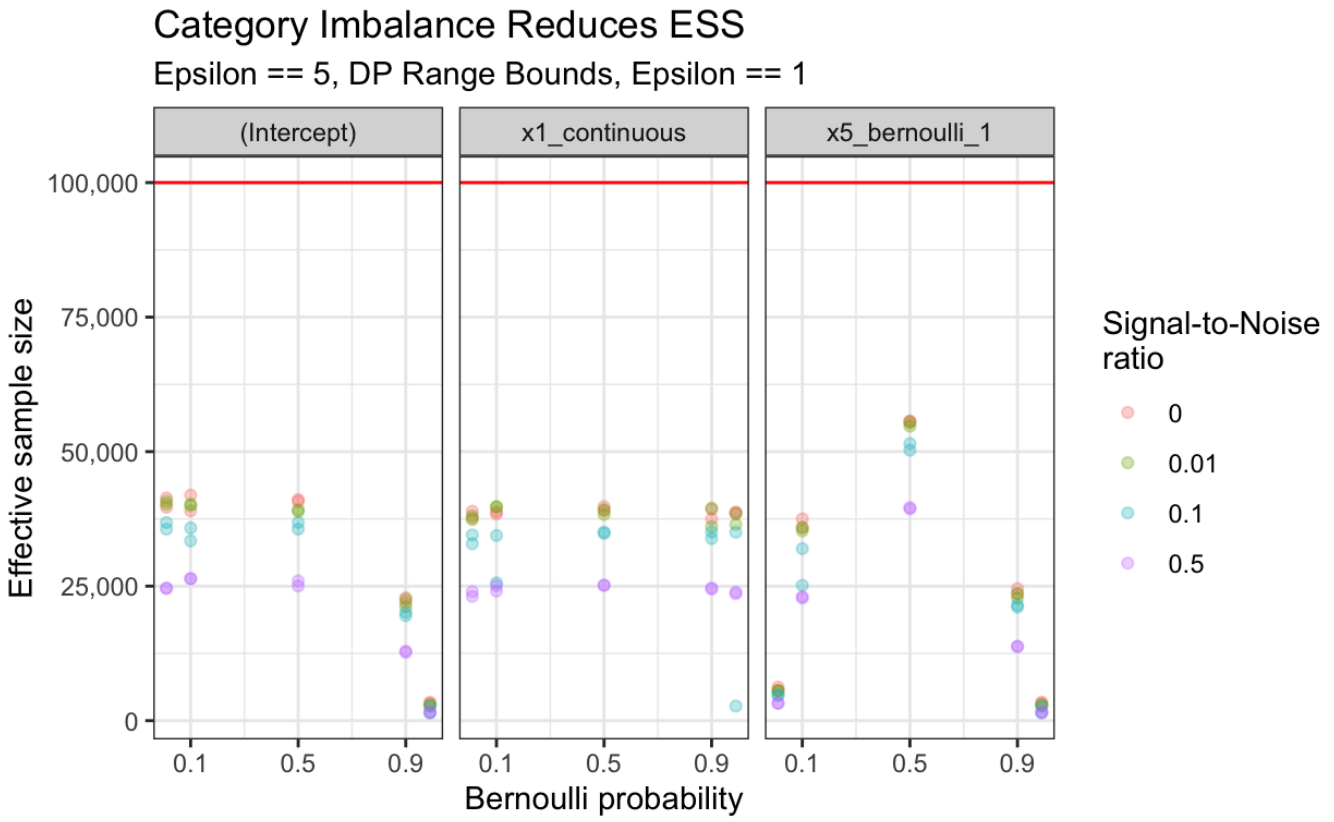
Omitting a Predictor

Bootstrap CIs, Epsilon == 1, DP Bounds Range, Analytic Gaussian Mechanism, SNR == 0.01

covariance	term	rep	coverage_probability	relative_length	bias	power	ess
Baseline							
NA	x1_continuous	100	98.0%	183.2%	0.0011	99.0%	37,842
Omitted predictor							
0.0	x1_continuous	20	100.0%	144.6%	0.0014	100.0%	47,875
0.5	x1_continuous	20	0.0%	144.9%	0.0000	100.0%	47,719
0.9	x1_continuous	20	0.0%	144.5%	0.0004	100.0%	47,928

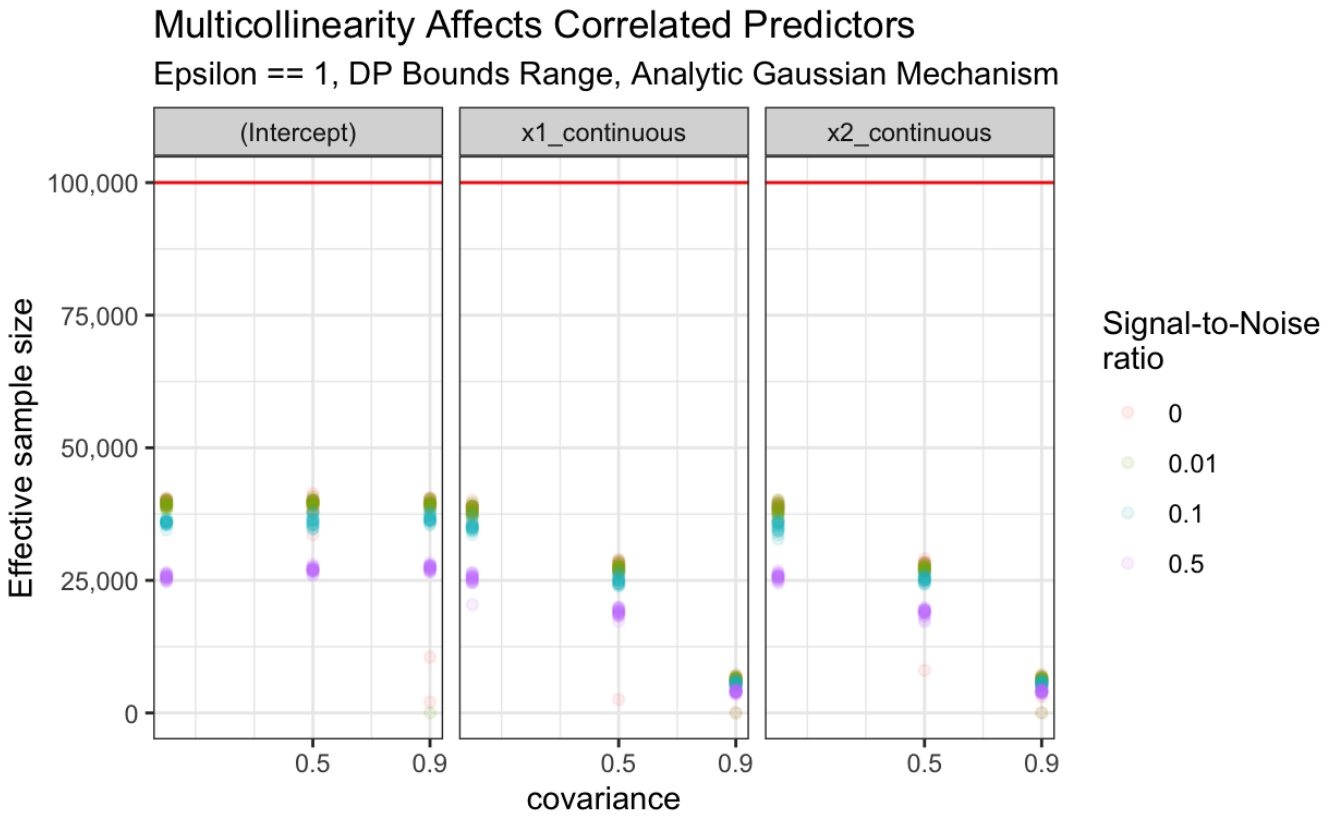
Takeaway #4: Violations of MLR assumptions affect results but not as much as #2 or #3.

- *Highly imbalanced categories reduces the effective sample size.*



Takeaway #4: Violations of MLR assumptions affect results but not as much as #2 or #3.

- *Increasing the covariance between x1_continuous and x2_continuous reduces the effective sample size.*



Discussion

Summary Takeaways

1. Context drives performance of DP regression.
2. Sample size, privacy-loss budget, and sensitivity matter.
3. Mechanisms and techniques for bounds matter but not as much as #2.
4. Violations of MLR assumptions matter but not as much as #2 or #3.
5. Framework is useful pre-testing and evaluation.

 awilliams@urban.org

 @awunderground

Appendix

What is a validation server?

Our working definition is a system that can:

- Accept submitted research programs (.R, .do, etc.)
- Automatically calculate and return privacy-preserving results
- Provide information about and enforce the “privacy budget” of released results for each researcher and across all users
- Empower the researcher to manage their privacy budget

Differential Privacy

Definition 1. ***Differential Privacy*** *(Dwork et al., 2006): A sanitization algorithm, $\mathcal{M}$, satisfies ϵ-DP if for all subsets $S \subseteq Range(\mathcal{M})$ and for all X, X' such that $d(X, X') = 1$,*

$$\frac{\Pr(\mathcal{M}(X) \in S)}{\Pr(\mathcal{M}(X') \in S)} \leq \exp(\epsilon) \tag{1}$$

where $\epsilon > 0$ is the privacy loss budget and $d(X, X') = 1$ represents the possible ways that X' differs from X by one record.

Definition 4. ***l_1-Global Sensitivity*** *(Dwork et al., 2006): For all X, X' such that $d(X, X') = 1$, the global sensitivity of a function M is*

$$\Delta_1(M) = \sup_{d(X,X')=1} \|M(X) - M(X')\|_1 \tag{4}$$

Baseline Simulation

$$\begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \text{Normal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_{X_1,X_2} \\ \sigma_{X_1,X_2} & \sigma_X^2 \end{pmatrix} \right), \quad i = 1, \ldots, n,$$

where $\sigma_{X_1,X_2} = 0$ and $\sigma_X^2 = 1$.

$$\begin{pmatrix} X_{i,3} \\ X_{i,4} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \text{Multinomial}(1, (\pi_a, \pi_b, \pi_c))$$

$$X_{i,5} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\pi_d), \qquad i = 1, \ldots, n.$$

where π_a=π_b=π_c=1/3 and π_d=π_e=1/2

Simulations

$$\sigma_{X_1} = \left(1 - \frac{\tau}{2}\right) + \frac{\tau \exp(X_1)}{1 + \exp(X_1)}$$

Privacy

- We need to know the sensitivity of $\boldsymbol{S}$ to create $\boldsymbol{H}$ for DP
- We place bounds on the continuous variables, $\vec{Y}, \overrightarrow{X_1}, \overrightarrow{X_2}$, to calculate the sensitivity
    1. Local bounds: use observed minimum and maximum for each variable
    2. DP range bounds: estimate noisy extreme percentiles (0.001, 0.999) to approximate the minimum and maximum for each variable
        1. Use the joint exponential mechanism for the percentiles
        2. $\epsilon = \frac{1}{3}$ for each variable
        3. Use 100 times the observed range as the prior

Previous Work

- "Disclosing Economists' Privacy Perspectives: A Survey of American Economic Association Members on Differential Privacy and Data Fitness for Use Standards" (Williams at al., 2024)*
  1. Empirical estimates of fitness for use based on sign match, significance match, absolute relative error, and CI ratio

*Presented at "Data Privacy Protection and the Conduct of Applied Research: Methods, Approaches and Their Consequences," Spring 2023 and accepted to HDSR special issue.