New Formal Privacy Methods for Business Populations

Webinar Presented by:

Margaret Beckom, *Economic Management Division*

William Sexton, *Tumult Labs*

Anthony Caruso, *Economic Statistical Methods Division*

United States® Census Bureau

June 9, 2023

What will be covered…

- Outreach Plans & Outstanding Research
- County Business Patterns (CBP) Background
- High-level Overview of Per-Record Differential Privacy (PRDP)
- Accuracy & Privacy
- Second-Stage Noise
- Parameter Selection
- Demonstration Tables

United States® Census Bureau

Outreach Plans & Outstanding Research

A Look into Planned Outreach Activities & Outstanding Research

United States® Census Bureau

Outreach Plans

- Internal presentations for:
  - Senior leadership
  - Data Stewardship Executive Policy Committee
  - Disclosure Review Board
  - Methodology & Standards Council

- Demonstration Tables
  - Two planned releases with Federal Register Notices
  - Webinars

United States® Census Bureau

Outstanding Research

- Protecting sample-based estimates
- Privacy-conserving approaches for firm counts
- Implications for benchmarking
- Privacy-protection algorithms for functions of the data other than sums
- Required privacy protections for product-level statistics

United States® Census Bureau

County Business Patterns

A Brief Overview on the Program and Current Disclosure Avoidance Methodology

United States® Census Bureau

Program Background

- Includes the following estimates
  - Counts
    - Establishments
  - Magnitude
    - Employment during the week of March 12
    - First quarter payroll
    - Annual payroll
- Data is useful for studying the economic activity in small areas
- Current methodology: multiplicative noise

United States® Census Bureau

A Look into Per-Record DP

How to handle heavy-tailed distributions

United States® Census Bureau

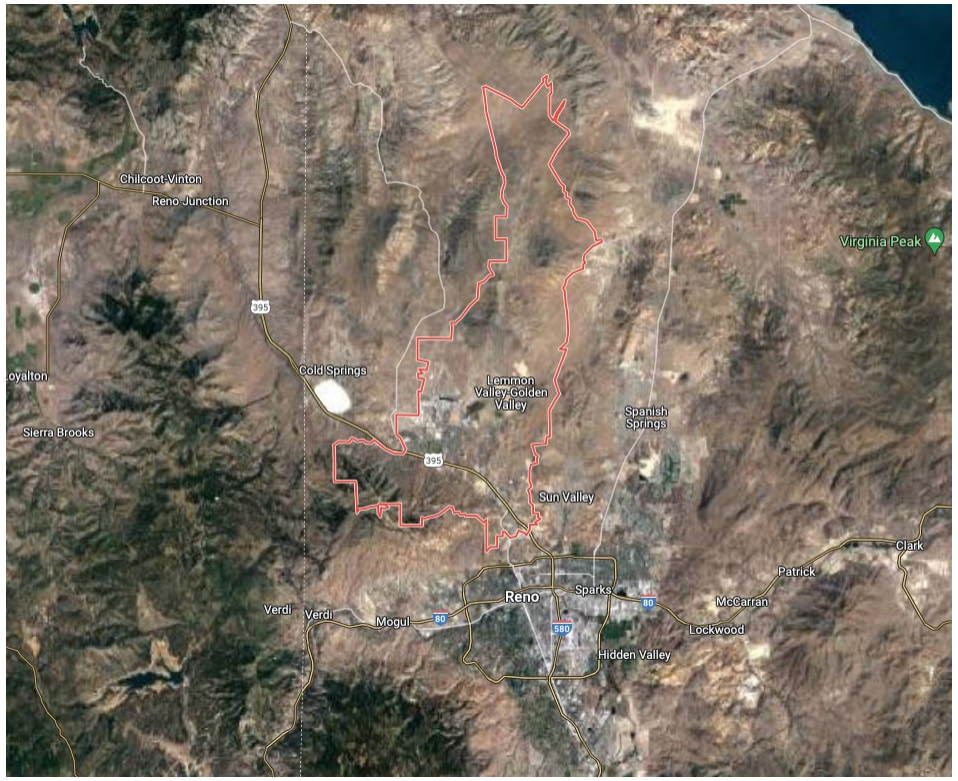
The Challenge & Takeaways

- Accurately release key economic indicators from heavy-tailed distributions with modernized privacy protection

- Differential privacy (DP) provides strong privacy protection but does not handle heavy-tailed distributions very well.
- “Per-Record” DP (PRDP) provides high data utility and formal privacy protection, but the privacy protection is not as strong as differential privacy.

United States® Census Bureau

The challenge of applying DP to the CBP: How many employees work in 89506?

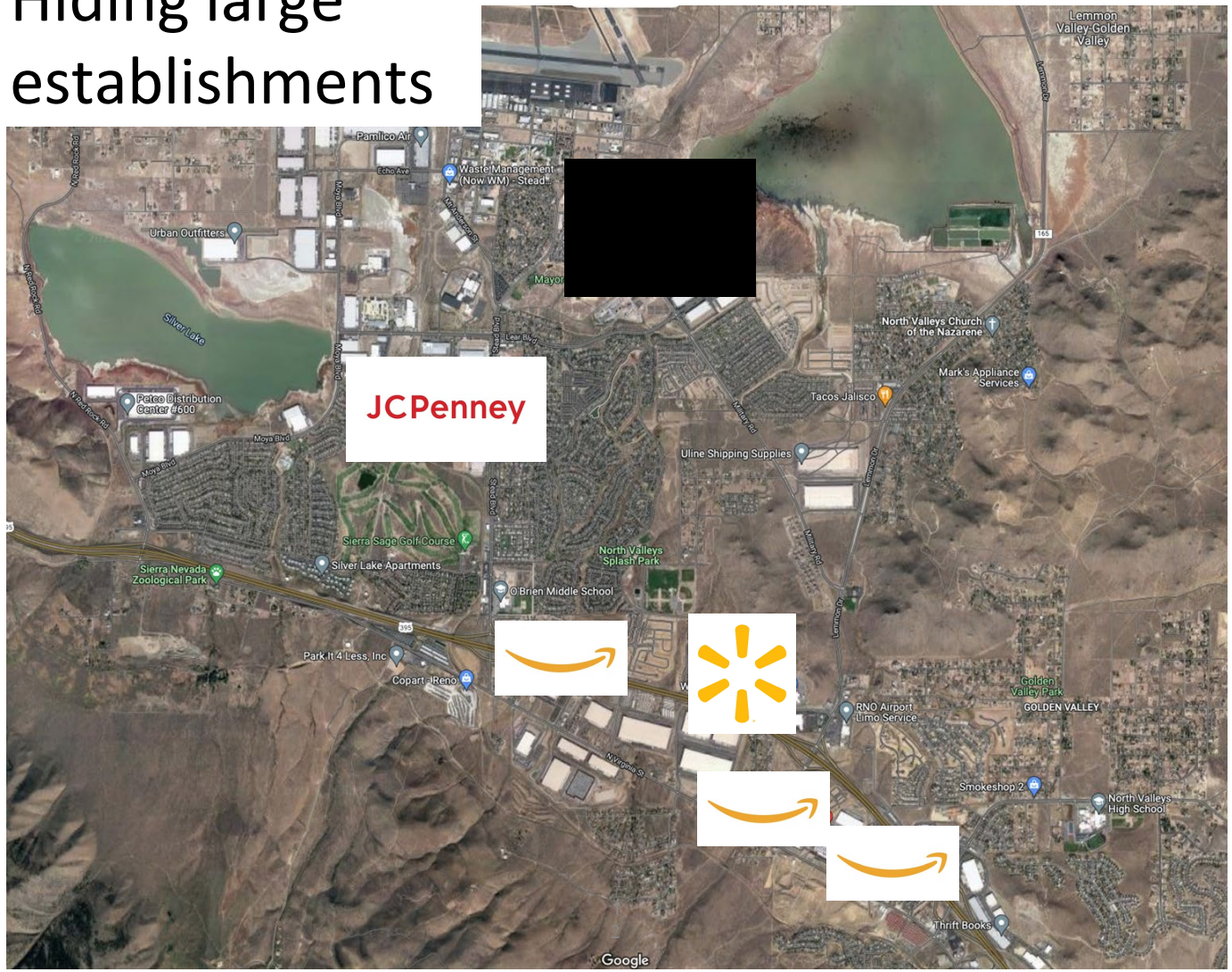


United States® Census Bureau

Hiding large establishments

United States® Census Bureau

Hiding large establishments



United States® Census Bureau

Overkill for small establishments

United States® Census Bureau

Solution: Design a formal privacy framework that provides “sliding” protection to establishments.

Protection against “fact-of-filing”: adversaries should not be able to easily infer whether an establishment is represented in the CBP dataset.

Protection against exact inference: adversaries should not be able to deterministically infer exact attributes about an establishment, such as employee size or payrolls.

Protection for firms: the privacy properties of firms (i.e., collections of related establishments) should be inherited from the privacy properties of their individual establishments.

“Sliding” establishment protection: Allow the privacy guarantees to vary by establishment. In particular, allow privacy guarantees to degrade as the influence of an establishment grows.

United States® Census Bureau

Hide small establishments?

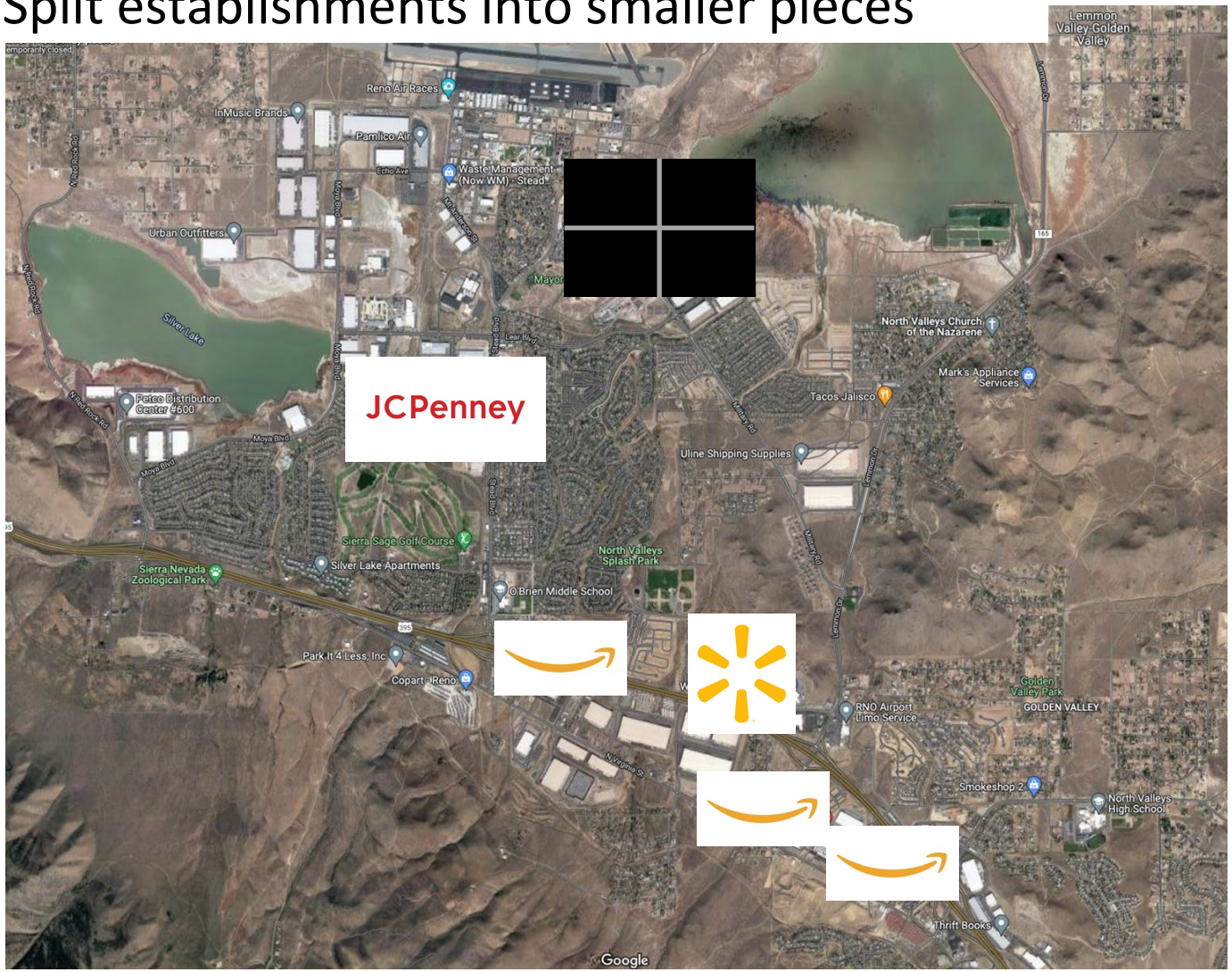
United States® Census Bureau

Fully hide small establishments

Hide large establishments?

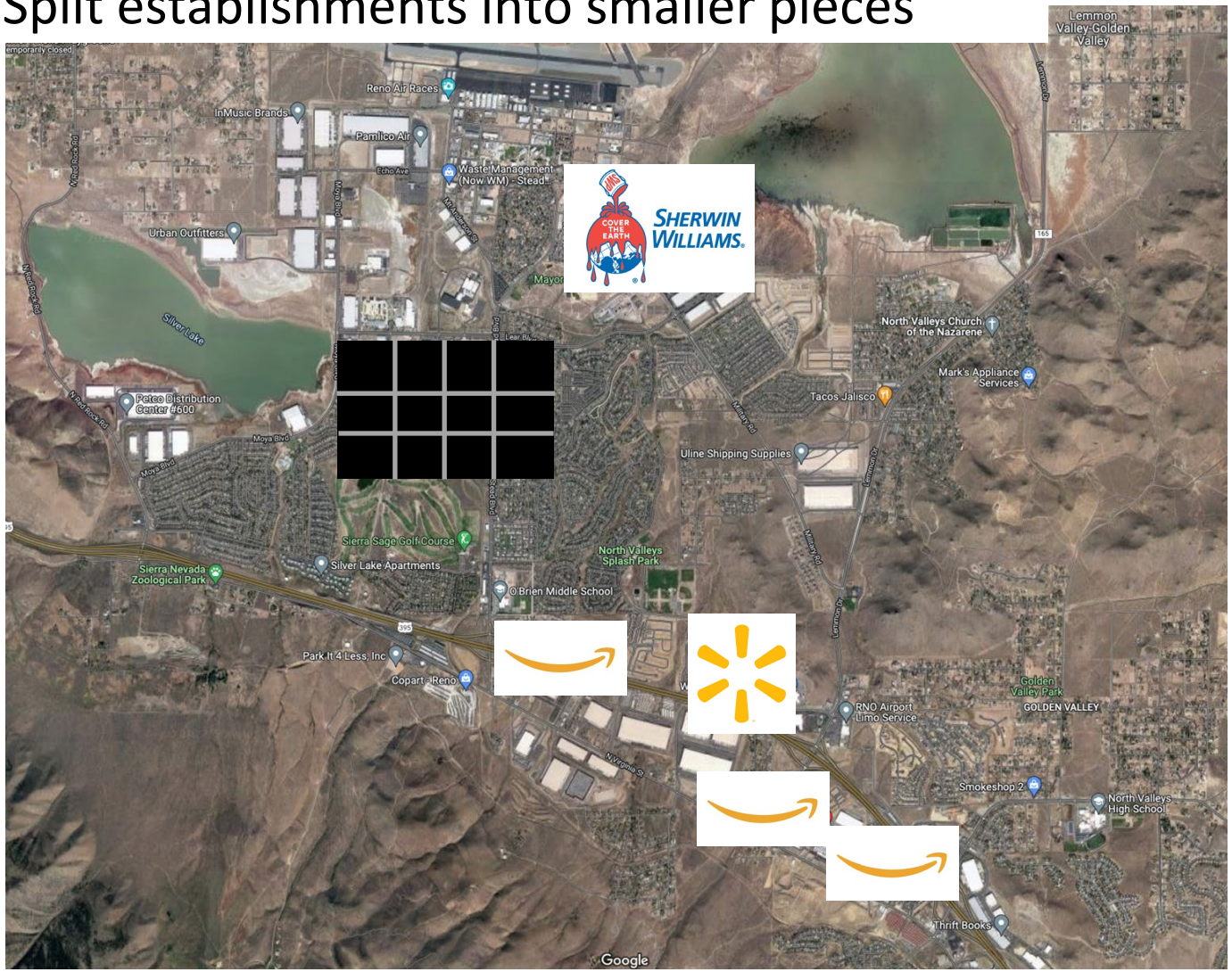
United States® Census Bureau

Split establishments into smaller pieces



United States® Census Bureau

Split establishments into smaller pieces



Group Privacy for ε-DP Algorithms

- ε-DP algorithm privacy guarantee extends to groups of size *k*

- When input databases differ by adding or removing up to *k* records:
  - Output distributions are bounded by *k* x ε

United States® Census Bureau

PRDP for Establishments

- When input databases differ by adding or removing a particular record, r
  - Output distributions are bounded by a function P(r)

- With establishment splitting A followed by DP mechanism M
  - $P(r) = |A(r)| \times \varepsilon$ where $|A(r)|$ is the number of pieces record r gets split into

United States® Census Bureau

Data Utility Comparison

No establishment splitting

- Bias-variance trade off
- Outliers drive noise requirements
- Outliers dominate a large share of the aggregations

Establishment splitting

- No bias
- Noise requirements can be calibrated based on smaller establishments

United States® Census Bureau

Accuracy and Privacy

Overview & Approach

United States® Census Bureau

Accuracy and Privacy

- **The tradeoff between accuracy and privacy remains.**

  Less privacy for establishments → More accurate estimates
  More privacy for establishments → Less accurate estimates

- Challenge: Find a balance where quality is preserved, and all establishments are adequately protected.

United States® Census Bureau

Median CV of County x 3-Digit NAICS Annual Payroll Estimates by Standard Deviation of Noise and Cell Size

Standard deviation of noise	All	Cell size (number of establishments)				
		a) 1-2	b) 3-9	c) 10-24	d) 25-99	e) 100+
10	0.00	0.03	0.01	0.00	0.00	0.00
20	0.01	0.07	0.02	0.00	0.00	0.00
50	0.02	0.17	0.04	0.01	0.00	0.00
100	0.04	0.35	0.08	0.02	0.01	0.00
200	0.09	0.69	0.16	0.04	0.01	0.00
500	0.22	>1.00	0.40	0.11	0.03	0.00
1,000	0.44	>1.00	0.80	0.22	0.06	0.01
2,000	0.88	>1.00	>1.00	0.44	0.13	0.02
5,000	>1.00	>1.00	>1.00	>1.00	0.31	0.04
10,000	>1.00	>1.00	>1.00	>1.00	0.63	0.09
20,000	>1.00	>1.00	>1.00	>1.00	>1.00	0.18
# of cells	187,446	53,264	61,799	32,569	26,901	12,913

United States® Census Bureau

Median CV of County x 3-Digit NAICS Annual Payroll Estimates by Standard Deviation of Noise and Cell Size

Standard deviation of noise	All	Cell size (number of establishments)				
		a) 1-2	b) 3-9	c) 10-24	d) 25-99	e) 100+
10	0.00	0.03	0.01	0.00	0.00	0.00
20	0.01	0.07	0.02	0.00	0.00	0.00
50	0.02	0.17	0.04	0.01	0.00	0.00
100	0.04	0.35	0.08	0.02	0.01	0.00
200	0.09	0.69	0.16	0.04	0.01	0.00
500	0.22	>1.00	0.40	0.11	0.03	0.00
1,000	0.44	>1.00	0.80	0.22	0.06	0.01
2,000	0.88	>1.00	>1.00	0.44	0.13	0.02
5,000	>1.00	>1.00	>1.00	>1.00	0.31	0.04
10,000	>1.00	>1.00	>1.00	>1.00	0.63	0.09
20,000	>1.00	>1.00	>1.00	>1.00	>1.00	0.18
# of cells	187,446	53,264	61,799	32,569	26,901	12,913

United States® Census Bureau

Ensuring Adequate Protection

- PRDP with quality-preserving parameters will provide adequate protection for most establishments' magnitude data.
- Very large establishments will not have adequate protection.

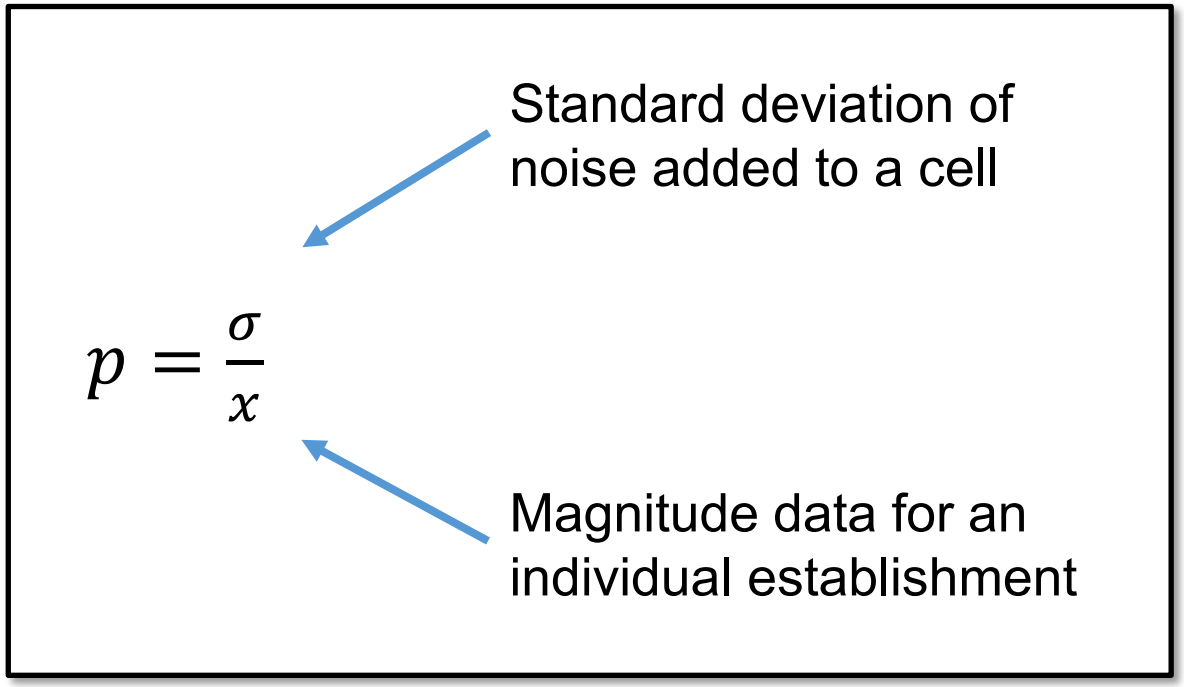
Establishment	Employment
#1	75
#2	150
#3	30,000
Total	**30,225**

$$\xrightarrow[\text{and } \rho = 0.1]{\text{PRDP with } \theta = 10}$$

30,200 ± 37 (90% CI)

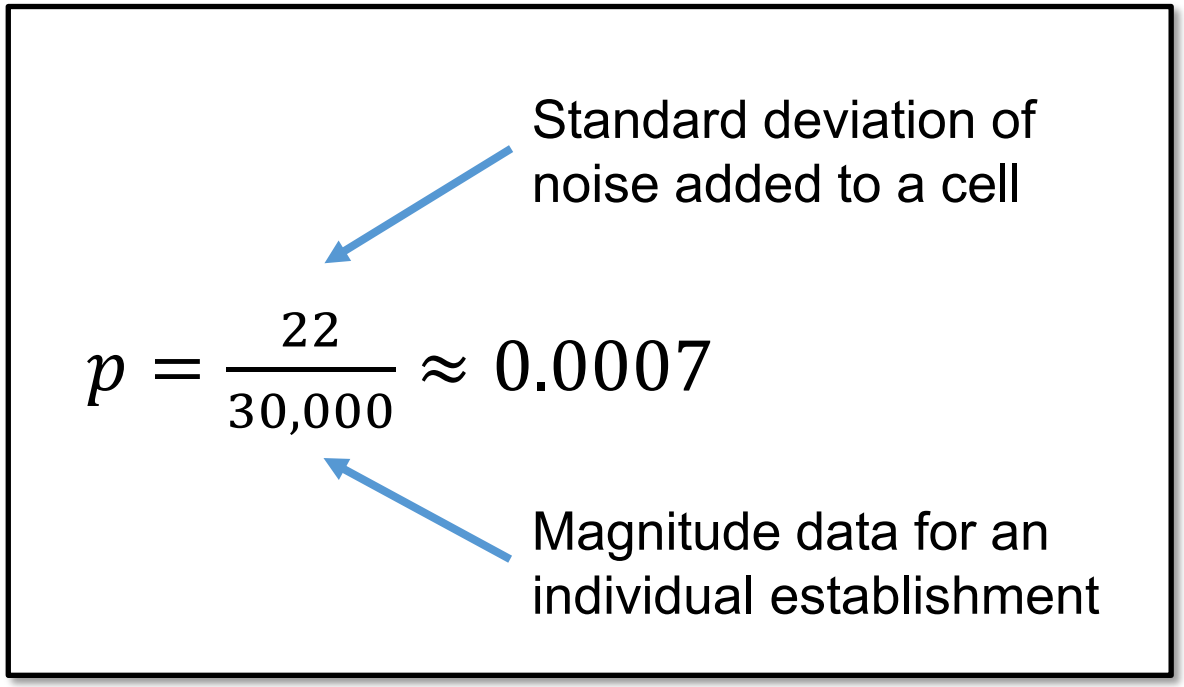
United States® Census Bureau

Relative Protection



United States® Census Bureau

Relative Protection



United States® Census Bureau

Proportion of Establishments (N=7,960,386) with Relative Protection* Meeting or Exceeding Selected Levels

Standard deviation of noise	p ≥ 0.001	p ≥ 0.01	p ≥ 0.02	p ≥ 0.05	p ≥ 0.10	p ≥ 0.15	p ≥ 0.20	p ≥ 1.00
10	0.987	0.877	0.792	0.633	0.489	0.402	0.346	0.111
20	0.994	0.932	0.877	0.758	0.633	0.550	0.489	0.191
50	0.998	0.972	0.945	0.877	0.792	0.728	0.676	0.346
100	0.999	0.987	0.972	0.932	0.877	0.831	0.792	0.489
200	1.000	0.994	0.987	0.965	0.932	0.903	0.877	0.633
500	1.000	0.998	0.996	0.987	0.972	0.958	0.945	0.792
1,000	1.000	0.999	0.998	0.994	0.987	0.980	0.972	0.877
2,000	1.000	1.000	0.999	0.998	0.994	0.991	0.987	0.932
5,000	1.000	1.000	1.000	0.999	0.998	0.997	0.996	0.972
10,000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.987
20,000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994

* Relative protection for annual payroll

United States® Census Bureau

Number of Establishments (N=7,960,386) with Relative Protection* Less than Selected Levels

Standard deviation of noise	p < 0.001	p < 0.01	p < 0.02	p < 0.05	p < 0.10	p < 0.15	p < 0.20	p < 1.00
10	103,250	982,779	1,655,356	2,920,852	4,067,289	4,761,364	5,205,203	7,074,254
20	45,312	542,070	982,779	1,926,842	2,920,852	3,585,832	4,067,289	6,439,113
50	14,304	221,237	441,006	982,779	1,655,356	2,168,765	2,578,852	5,205,203
100	5,651	103,250	221,237	542,070	982,779	1,346,866	1,655,356	4,067,289
200	2,169	45,312	103,250	278,455	542,070	774,739	982,779	2,920,852
500	502	14,304	34,458	103,250	221,237	334,079	441,006	1,655,356
1,000	147	5,651	14,304	45,312	103,250	162,836	221,237	982,779
2,000	42	2,169	5,651	18,995	45,312	73,773	103,250	542,070
5,000	6	502	1,549	5,651	14,304	23,987	34,458	221,237
10,000	2	147	502	2,169	5,651	9,766	14,304	103,250
20,000	1	42	147	737	2,169	3,785	5,651	45,312

* Relative protection for annual payroll

United States® Census Bureau

Number of Establishments (N=7,960,386) with Relative Protection* Less than Selected Levels

Standard deviation of noise	p < 0.001	p < 0.01	p < 0.02	p < 0.05	p < 0.10	p < 0.15	p < 0.20	p < 1.00
10	103,250	982,779	1,655,356	2,920,852	4,067,289	4,761,364	5,205,203	7,074,254
20	45,312	542,070	982,779	1,926,842	2,920,852	3,585,832	4,067,289	6,439,113
50	14,304	221,237	441,006	982,779	1,655,356	2,168,765	2,578,852	5,205,203
100	5,651	103,250	221,237	542,070	982,779	1,346,866	1,655,356	4,067,289
200	2,169	45,312	103,250	278,455	542,070	774,739	982,779	2,920,852
500	502	14,304	34,458	103,250	221,237	334,079	441,006	1,655,356
1,000	147	5,651	14,304	45,312	103,250	162,836	221,237	982,779
2,000	42	2,169	5,651	18,995	45,312	73,773	103,250	542,070
5,000	6	502	1,549	5,651	14,304	23,987	34,458	221,237
10,000	2	147	502	2,169	5,651	9,766	14,304	103,250
20,000	1	42	147	737	2,169	3,785	5,651	45,312

* Relative protection for annual payroll

United States® Census Bureau

Number of Establishments (N=7,960,386) with Relative Protection* Less than Selected Levels

Standard deviation of noise	p < 0.001	p < 0.01	p < 0.02	p < 0.05	p < 0.10	p < 0.15	p < 0.20	p < 1.00
10	103,250	982,779	1,655,356	2,920,852	4,067,289	4,761,364	5,205,203	7,074,254
20	45,312	542,070	982,779	1,926,842	2,920,852	3,585,832	4,067,289	6,439,113
50	14,304	221,237	441,006	982,779	1,655,356	2,168,765	2,578,852	5,205,203
100	5,651	103,250	221,237	542,070	982,779	1,346,866	1,655,356	4,067,289
200	2,169	45,312	103,250	278,455	542,070	774,739	982,779	2,920,852
500	502	14,304	34,458	103,250	221,237	334,079	441,006	1,655,356
1,000	147	5,651	14,304	45,312	103,250	162,836	221,237	982,779
2,000	42	2,169	5,651	18,995	45,312	73,773	103,250	542,070
5,000	6	502	1,549	5,651	14,304	23,987	34,458	221,237
10,000	2	147	502	2,169	5,651	9,766	14,304	103,250
20,000	1	42	147	737	2,169	3,785	5,651	45,312

* Relative protection for annual payroll

United States® Census Bureau

Number of Establishments (N=7,960,386) with Relative Protection* Less than Selected Levels

Standard deviation of noise	p < 0.001	p < 0.01	p < 0.02	p < 0.05	p < 0.10	p < 0.15	p < 0.20	p < 1.00
10	103,250	982,779	1,655,356	2,920,852	4,067,289	4,761,364	5,205,203	7,074,254
20	45,312	542,070	982,779	1,926,842	2,920,852	3,585,832	4,067,289	6,439,113
50	14,304	221,237	441,006	982,779	1,655,356	2,168,765	2,578,852	5,205,203
100	5,651	103,250	221,237	542,070	982,779	1,346,866	1,655,356	4,067,289
200	2,169	45,312	103,250	278,455	542,070	774,739	982,779	2,920,852
500	502	14,304	34,458	103,250	221,237	334,079	441,006	1,655,356
1,000	147	5,651	14,304	45,312	103,250	162,836	221,237	982,779
2,000	42	2,169	5,651	18,995	45,312	73,773	103,250	542,070
5,000	6	502	1,549	5,651	14,304	23,987	34,458	221,237
10,000	2	147	502	2,169	5,651	9,766	14,304	103,250
20,000	1	42	147	737	2,169	3,785	5,651	45,312

* Relative protection for annual payroll

United States® Census Bureau

Second-Stage Noise

- Additional noise to increase relative protection
- Added post-PRDP
- Scaled to noisy sums
  - ***NOT*** the largest establishment in each cell
- Not formally private

United States® Census Bureau

Second-Stage Noise

Establishment	Employment
#1	75
#2	150
#3	30,000
Total	**30,225**

$$\xrightarrow[\text{and } \rho = 0.1]{\text{PRDP with } \theta = 10}$$

30,200 ± 37 (90% CI)

United States® Census Bureau

Second-Stage Noise

Establishment	Employment
#1	75
#2	150
#3	30,000
Total	**30,225**

PRDP with $\theta = 10$ and $\rho = 0.1$ → **30,200 ± 37 (90% CI)**

Additional noise to ensure $p \approx 0.10$

28,500 ± 4,688 (90% CI)

United States® Census Bureau

Parameter Selection

A brief look at the parameter tuning approach

United States® Census Bureau

Privacy Loss Budget (ρ) for PRDP

- Quality Target: 95% of cells with at least ***x*** establishments have a CV of 0.10 or less at each tabulation level

x	Establishments	Annual Payroll (θ = $100,000)	First Quarter Payroll (θ = $25,000)	Employment (θ = 4)	Total Privacy Loss Budget (ρ)
1	3160.55	7597.712	7219.692	5019.129	22997.083
10	30.821	7.666	9.537	19.804	67.828
25	5.066	1.381	1.702	3.909	12.058
100	0.342	0.133	0.158	0.375	1.008

United States® Census Bureau

Privacy Loss Budget (ρ) for PRDP

- Quality Target: 95% of cells with at least **25** establishments have a CV of 0.10 or less at each tabulation level

x	Establishments	Annual Payroll (θ = $100,000)	First Quarter Payroll (θ = $25,000)	Employment (θ = 4)	Total Privacy Loss Budget (ρ)
1	3160.55	7597.712	7219.692	5019.129	22997.083
10	30.821	7.666	9.537	19.804	67.828
25	5.066	1.381	1.702	3.909	12.058
100	0.342	0.133	0.158	0.375	1.008

United States® Census Bureau

Second-Stage Noise

- Cells have a minimum CV via second-stage noise:

$$CV_{MINIMUM} = \max\left(0, \min\left(0.25, 0.25 - \frac{\#\text{ noisy establishments}}{100}\right)\right)$$

# of Noisy Establishments	Minimum CV
≤ 0	25%
1	24%
5	20%
15	10%
25+	0%

United States® Census Bureau

Demonstration Tables

- All CBP tabulation levels
  - Non-noisy and noisy values
  - CVs/variances
- Summary table:
  - Mean absolute error (MAE)
  - Mean absolute percent error (MAPE)
  - Median coefficient of variation
  - 95th percentile coefficient of variation
  - A comparison to noise ranges in the published 2019 CBP tables (for rows with at least 3 establishments)

United States® Census Bureau

Contact Us

- Margaret Beckom
  Dissemination Standards Branch
  Economic Management Division

  Email: margaret.m.beckom@census.gov
  with the subject “CBP Disclosure Feedback”

  Phone: 301-763-7522

United States® Census Bureau